

SEARCH REQUEST FORM

Access DB# _____

Scientific and Technical Information Center

Requester's Full Name: _____ Examiner #: _____ Date: _____
 Art Unit: _____ Phone Number 30 _____ Serial Number: _____
 Mail Box and Bldg/Room Location: _____ Results Format Preferred (circle): PAPER DISK E-MAIL

If more than one search is submitted, please prioritize searches in order of need.

 Please provide a detailed statement of the search topic, and describe as specifically as possible the subject matter to be searched. Include the elected species or structures, keywords, synonyms, acronyms, and registry numbers, and combine with the concept or utility of the invention. Define any terms that may have a special meaning. Give examples or relevant citations, authors, etc. if known. Please attach a copy of the cover sheet, pertinent claims, and abstract.

Title of Invention: _____

Inventors (please provide full names): _____

Earliest Priority Filing Date: _____

**For Sequence Searches Only* Please include all pertinent information (parent, child, divisional, or issued patent numbers) along with the appropriate serial number.*

STAFF USE ONLY

| | | |
|--|--------------------------|-----------------------------------|
| Searcher: _____ | Type of Search | Vendors and cost where applicable |
| Searcher Phone #: _____ | NA Sequence (#) <u>3</u> | STN _____ |
| Searcher Location: _____ | AA Sequence (#) <u>1</u> | Dialog _____ |
| Date Searcher Picked Up: <u>10/17/01</u> | Structure (#) _____ | Questel/Orbi: _____ |
| Date Completed: <u>10/18/01</u> | Bibliographic _____ | Dr Link _____ |
| Searcher Prep & Review Time _____ | Litigation _____ | Lexis/Nexis _____ |
| Clencal Prep Time _____ | Fulltext _____ | Sequence Systems <u>\$3</u> |
| Online Time: _____ | Patent Family _____ | WWW/Internet _____ |
| | Other _____ | Other (specify) _____ |

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 17, 2001, 14:43:25 ; Search time 9769.47 Seconds

(without alignments)
5069.644 Million cell updates/sec

Title: US-09-502-426-1_COPY_1_3202

Perfect score: 3202

Sequence: 1 atgtgggtatattatgttg.....gagagagagaactagcttc 3202

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 1344157 seqs, 7733874588 residues

Total number of hits satisfying chosen parameters: 2688314

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : GenEmbl:*
1: gb_ba1:*
2: gb_ba2:*
3: gb_ba3:*
4: gb_in1:*
5: gb_in2:*
6: gb_in3:*
7: gb_om:*
8: gb_ov:*
9: gb_pat1:*
10: gb_pat2:*
11: gb_ph:*
12: gb_pl1:*
13: gb_pl2:*
14: gb_pl3:*
15: gb_pl4:*
16: em_ba1:*
17: em_ba2:*
18: em_fun:*
19: em_htgo_hum:*
20: em_htgo_inv:*
21: em_htgo_rod:*
22: em_htg_hum1:*
23: em_htg_hum2:*
24: em_htg_hum3:*
25: em_htg_hum4:*
26: em_htg_hum5:*
27: em_htg_hum6:*
28: em_htg_hum7:*
29: em_htg_hum8:*
30: em_htg_inv1:*
31: em_htg_inv2:*
32: em_htg_other:*
33: em_htg_rod:*
34: em_hum1:*
35: em_hum2:*
36: em_hum3:*
37: em_hum4:*
38: em_hum5:*
39: em_hum6:*
40: em_hum7:*
41: em_in:*
42: em_om:*
43: em_or:*

44: em_ov:*
45: em_pat:*
46: em_ph:*
47: em_pl:*
48: em_ro:*
49: em_sts:*
50: em_sy:*
51: em_un:*
52: em_v1:*
53: gb_sts1:*
54: gb_sts2:*
55: gb_sts3:*
56: gb_sy:*
57: gb_un:*
58: gb_v11:*
59: gb_v12:*
60: gb_htg1:*
61: gb_htg2:*
62: gb_htg3:*
63: gb_htg4:*
64: gb_htg5:*
65: gb_htg6:*
66: gb_htg7:*
67: gb_htg8:*
68: gb_htg9:*
69: gb_htg10:*
70: gb_htg11:*
71: gb_htg12:*
72: gb_htg13:*
73: gb_htg14:*
74: gb_htg15:*
75: gb_htg16:*
76: gb_htg17:*
77: gb_htg18:*
78: gb_htg19:*
79: gb_htg20:*
80: gb_htg21:*
81: gb_htg22:*
82: gb_htg23:*
83: gb_htg24:*
84: gb_htg25:*
85: gb_pr1:*
86: gb_pr2:*
87: gb_pr3:*
88: gb_pr4:*
89: gb_pr5:*
90: gb_pr6:*
91: gb_pr7:*
92: gb_pr8:*
93: gb_pr9:*
94: gb_ro1:*
95: gb_ro2:*
96: gb_in4:*
97: gb_pr10:*
98: em_ba3:*

Pred. NO. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Match | Length | DB | ID | Description |
|------------|--------|-------|--------|----|----------|--------------------|
| C 1 | 3134.2 | 97.9 | 84196 | 14 | ATT3A5 | AL132979 Arabidops |
| 2 | 1132 | 35.4 | 4818 | 12 | AF044216 | AF044216 Arabidops |
| 3 | 133.2 | 4.2 | 104992 | 60 | AC005504 | AC005504 Plasmodu |
| 4 | 133.2 | 4.2 | 169546 | 60 | AC004157 | AC004157 Plasmodu |
| C 5 | 130.8 | 4.1 | 104992 | 60 | AC005504 | AC005504 Plasmodu |
| C 6 | 130.8 | 4.1 | 169546 | 60 | AC004157 | AC004157 Plasmodu |
| 7 | 129.4 | 4.0 | 14867 | 4 | AE001398 | AE001398 Plasmodu |
| 8 | 127.8 | 4.0 | 86827 | 96 | PFMAL3P5 | AL034556 Plasmodu |

```

source
1. 84196
/organism="Arabidopsis thaliana"
/variety="Columbia"
/db_xref="taxon:3702"
/chromosome="3"
2983..3338
LTR
/note="355 bp LTR"
3339..7564
misc_feature
/note="LTR-transposon"
7565..7922
5'
/note="367bp LTR"
12339..13197
gene
/note="33A5.10"
join(12339..12665,12742..12803,12888..13055,13125..13197)
/note="33A5.10"
/note="similarly to cyclin-dependent kinase inhibitor (CKI), Arabidopsis thaliana, EMBL:ATU94772_1"
(codon_start=1
/product="putative protein"
/protein_id="CAB62432.1"
/db_xref="GI:6561966"
/translation="MAARRRRRDVVEENGVTITTVKRRKKEEVDLVESRIISPCV
QATINNGIVATNSAGASSETSVIVRRRDPVEEQCIIEEDSSVSCSTSEESKRR
IEFVLEENNDDRETFETSWIYDLNKSSEEMNDDSSVAIVEDESRRRLKSLHETV
KEALEEDFEQVAKERDLRNLKLECSMKYNFDEKDEPLGGRYEMWKLP"
12339..12665
exon
/note="33A5.10"
/number=1
12666..12741
intron
/note="33A5.10"
/number=1
12742..12803
exon
/note="33A5.10"
/number=2
12804..12887
intron
/note="33A5.10"
/number=2

```

| Feature | Gene | Protein |
|---------|---|--|
| intron | 13056..13124 /gene="r3A5.10" | 13056..13124 /number=3 |
| exon | 13125..13197 /gene="r3A5.10" | 13125..13197 /number=3 |
| exon | 15858..16470 /gene="r3A5.20" | 15858..16470 /number=1 |
| gene | complement(join(15858..15901,16014..16470)) /note="similarity to predicted proteins, Arabidopsis thaliana" | complement(15858..15901) /number=1 |
| CDS | complement(join(15858..15901,16014..16470)) /note="similarity to predicted proteins, Arabidopsis thaliana" | complement(join(15858..15901,16014..16470)) /number=1 |
| intron | 16014..16470 /gene="r3A5.30" | 16014..16470 /number=2 |
| exon | 16471..16561 /gene="r3A5.30" | 16471..16561 /number=2 |
| exon | 16562..16652 /gene="r3A5.30" | 16562..16652 /number=2 |
| CDS | complement(16303..19931) /codon_start=1 /product="scarecrow-like 7 (SCL7)" /protein_id="CAB62434.1" | complement(16303..19931) /codon_start=1 /product="scarecrow-like 7 (SCL7)" /protein_id="CAB62434.1" |

| | | | |
|----|-------|--|-------|
| OY | 661 | tgaaataatcatcaagaagaaataagagatgcacaaaccaaaggtcttgynaataatgctccct | 720 |
| Db | 31540 | tgaaataatattatcaaaagagaataagagatgacaaacaaagggttggaataatggccct | 31481 |
| OY | 721 | gccagcttctctcaacaataatcgaacctatttgaattctcggatattcggttaaa | 780 |
| Db | 31480 | gccaccttctctcaacaataatcgaacctatttgaattctcggatattcggttaaa | 31421 |
| OY | 781 | ttctgcaataagcattgctgaanaataattattgtttaagctgaatccaatattgttcca | 840 |
| Db | 31420 | tttggcgatracagattgtgnaaaatattttttattgttagctgaattccaatattgttcca | 31361 |
| OY | 841 | ggatctgcataactctctgtttaagaacataattgtctctctcttctgttctctct | 900 |
| Db | 31360 | ggatatttgcatatattctgttttaagaacataattgtctctctcttctgttctctct | 31301 |
| OY | 901 | taacatatataatcgcggatataatgataaacaatgatatataccaacaacattgctgg | 960 |
| Db | 31300 | taacatatataatgctggatataatgataaacaatgatatataccaacaacattgctgg | 31241 |
| OY | 961 | acctatttgaataaactttctcgaacatttcgggaacactgactgcaccttaaaata | 1020 |
| Db | 31240 | acctatttgaataaactttctcgaacatttcgggaacactgactgcaccttaaaata | 31182 |
| OY | 1021 | cgatttcaacgctcactagcttgagaattactagcaataaagcaataaaggacccgttcaagc | 1080 |
| Db | 31181 | cgatttcaacgctcactagcttgagaattactagcaataaagcaataaaggacccgttcaagc | 31122 |
| OY | 1081 | tatttatatacaaggcttacaacactgaaataagcttgaaatcccttagaaaaatttggaaata | 1140 |
| Db | 31121 | tatttatatacaaggcttacaacactgaaataagcttgaaatcccttagaaaaatttggaaata | 31062 |
| OY | 1141 | ccggtctgatatgataatagatttagtgtaaacaaatgataatgataatgataatgata | 1200 |
| Db | 31061 | ccggtctgatatgataatagatttagtgtaaacaaatgataatgataatgataatgata | 31002 |
| OY | 1201 | acataatacaatctccttacaagaaaaaacaacttaagagaagcttaacataatcatalat | 1260 |
| Db | 31001 | acataatacaatctccttacaagaaaaaacaacttaagagaagcttaacataatcatalat | 30942 |
| OY | 1261 | gggtatgcatacccttacaagatagctataactaagagactcaagaatagtatagtatgctc | 1320 |
| Db | 30941 | gggtatgcatacccttacaagatagctataactaagagactcaagaatagtatagtatgctc | 30882 |
| OY | 1321 | gataaatgaattcacacacgcgcgtgataaataattatggaacgcgatgttcaacatcgca | 1380 |
| Db | 30881 | gataaatgaattcacacacgcgcgcgtgataaataattatggaacgcgatgttcaacatcgca | 30822 |
| OY | 1381 | ataatcatcttggcttgctoaacaataaaacaaacaaagaagaaaaagaaacgaatttt | 1440 |
| Db | 30821 | ataatcatcttggcttgctoaacaataaaacaaacaaagaagaaaaagaaacgaatttt | 30762 |
| OY | 1441 | cttggatcccatcttaagatcctcaaaatgcatagatcctttgggttacagtttcgaatcc | 1500 |
| Db | 30761 | cttggatcccatcttaagatcctcaaaatgcatagatcctttgggttacagtttcgaatcc | 30702 |
| OY | 1501 | cttaagaacgcttgaacacatcgcacattttaaattgtcttctttaaatacatcttaaat | 1560 |
| Db | 30701 | cttaagaacgcttgaacacatcgcacattttaaattgtcttctttaaatacatctttaa | 30642 |
| OY | 1561 | atttatcttgatggaatttaataagacgaactctgtaacatacaatatattatctag | 1620 |
| Db | 30641 | atttatcttgatggaatttaataagacgaactctgtaacatacaatatattatctag | 30582 |
| OY | 1621 | atactagatagtgatattatccaataataacttggatggtttaaacttaactctgttct | 1680 |
| Db | 30581 | atactagatagtgatattatccaataataacttggatggtttaaacttaactctgttct | 30522 |
| OY | 1681 | ttctacggtataaataataatcatcgcgggttaaaaaaggttcttctattcttcgcgaatgc | 1740 |
| Db | 30521 | ttctacggtataaataataatcatcgcgggttaaaaaaggttcttctattcttcgcgaatgc | 30463 |

| | | | |
|----|-------|--|-------|
| QY | 1741 | atgaagataaacctaatagtaccttaattcttcttgaaaagttaacccttaactcatabagatc | 1800 |
| Db | 30462 | ATGAAGATATAAACCTAATGACTTAATTTTGTGAAATATGAAACCTTTACTCATAGATT | 30400 |
| QY | 1801 | aattacgatagttttctgttgccaataagaacgcccctcacaacgttgatgcaattt | 18600 |
| Db | 30402 | AATTACCGTATGTTTTTGTGGCCATATGACACCCCTCAACAAGTGTAGTCAATTTTTT | 30340 |
| QY | 1861 | tcctgcaaatataaatttaggaatccaatcgtcactacataagaagaacagctgataat | 19200 |
| Db | 30342 | TCTGCAAAATATTAAATTAGGAATTCAAATGGTACTATCATATAGAAACAGCTGATTT | 30280 |
| QY | 1921 | acatttaatttaagaacaaaattcttgaaaaagtataaattcctaacaatataat | 19800 |
| Db | 30282 | ACATTTAATTAAAGACAAAATTTTGTGAAAATGTTAATAATTTCTACAAATATTATTA | 30222 |
| QY | 1981 | aattatagcctaataagatcttcccaatgtcttaaaatatttttttatattagta | 20400 |
| Db | 30222 | AATATAGCCTATTAATAGTATTTCTATGCTCTTAATAATTTTTTTTATATATTAGTA | 3016 |
| QY | 2041 | taaatacatcttagaaccataatagatgtgtgaattcaaatcctccatataattctg | 21000 |
| Db | 30162 | TAAATACATTATGAAACCAATATATAGTTGGTGATTCAAATATCTCATTAATTTTTT | 3010 |
| QY | 2101 | aaactacaatatataatcttaagtaacaatagcaatagaaagtctccaaaaaat | 21600 |
| Db | 30102 | AAACTACAAATTTATTAATTTTATAGTCATATACAAATGCTATGAAAGTTCCAAAAAAT | 3004 |
| QY | 2161 | tgttaacaagaacttccaatttttttttatggacaagaataacagatagaaaa | 22200 |
| Db | 30042 | TTGTTAACGAACCTTCACAAATTTTTTTTTTTATGGAACAAAGAAATATACGATGAAA | 2998 |
| QY | 2221 | ctaatctgtcttggaacggaagtagtaatacatcaatgaagcaattctaaaaatata | 22800 |
| Db | 29982 | CTAATTGTGTGGAATAGGAATAGTAATATATCATTAATAGCAAAATTTTAAAAATATATA | 2992 |
| QY | 2281 | agcctaacgcgtccaagtatgtatctagtatgtatataatagcatgtgtgcgt | 23400 |
| Db | 29922 | AGCCATATCGCGGCTCAAAAGATGTTATCTATAGTGGGTATTAATAATGATGGTGCAT | 2986 |
| QY | 2341 | tcgaatttggaacaacaatgaaaaacgaatataaatataacttaataataataaat | 24000 |
| Db | 29862 | TCAGAAATGGGACACAAATATAAAACGCAATTAATAATTTAACTTTAAATAATTAATAAT | 2980 |
| QY | 2401 | cttgataaattgttcttctgactatctgaaggggcaaaaaagacaaatgccaaagtctac | 24600 |
| Db | 29802 | TTGAGTAAATGTGTTTCTCTCAGTATTGAGGGGCAAAAAAGACAAATGCCAAAGTGTAC | 2974 |
| QY | 2461 | gggttgactgtccagatctcgtgataatacctaabaactcgtctcttgacgcgaagctcgt | 25200 |
| Db | 29742 | GGGTTTGACTGTCCAGTTCGGTAATATCATTAATACCTGTCTTTGACCGGACGCTGTG | 2968 |
| QY | 2521 | taaggagctcttcgaactttcaactgtcttaccocactcggagcccaacccttccca | 25800 |
| Db | 29682 | TAGGGGTCCTTGACATTTTTCACGTCTTACCCCTACTCGTAGCCCAACCCCTTTTCCCA | 2962 |
| QY | 2581 | catcctaagggttaattcttgnaatcccaatttaaacggaatgagacgttacggactctc | 26400 |
| Db | 29622 | TATCTTAAGGTTAATTTTGGAAATTCCTCAATTTAAACCGATTGAGACGCTACCGGACTTCC | 2956 |
| QY | 2641 | tggaattctgtcgtgagcattatccaanaatctatgaacgaatgggttatataat | 27000 |
| Db | 29562 | TGGGATTTCTGCTGAGCATTTATTCAAAATTTATGACAGATGGGTTATTAATTTTAA | 2950 |
| QY | 2701 | aactcaaacacttgatcagataaaaattcatalaaacactttacgatgatctcgatcagatc | 27600 |
| Db | 29502 | AACTCACAACTTGATGATGATTAATAATTTTCATTAACACTTTTATGATGATGATTCGTACGATCT | 2944 |
| QY | 2761 | atcctaagactttttttcttccacacggttgatgaagaatagatataatgaacagag | 28200 |
| Db | 29442 | ATCTAATGACTTTTTTTTTTTCTACCAAGGTGATGAAGTTATATGACTATTATTCACGAG | 2938 |
| QY | 2821 | acaattatcatagatatataccaatcaaccgatatatttatataataaattgcgcttaaa | 2880 |

```

|||||
Db 29382 ACAATTGATTATAGATATATTCATTATTCATGATATTTATGATATTAATAGCTGTAA 29323
Oy 2881 cttttccacacgcagccttctgcacatttgcttttaatttaagaatttaataa 2940
Db 29322 CTAATTCACATCGCGAGCTTCTGCACTTTGTTTAAATTAAGAGTTTAAATATA 29263
Oy 2941 aagtaataaagagacatacagagcaacaaagtaatacagacgagaaacaaagca 3000
Db 29262 AAGTATTAAAGAGAGATACAGGACCAAAAGTATACACGAGAAACAAAGCCA 29203
Oy 3001 tgaagctcatggtta-gttaagcttaataagaagatttataaatttaagcat 3059
Db 29202 TGAAGTCTCATGTTGTTAGTTTAAGCTTAATAAGAGATTATTAATTTAATGACGAT 29143
Oy 3060 gataacatataatttctgcacttcttaaacccctcttaacaacagaagctccctt 3119
Db 29142 GATTAACAATTATATTTTCTGACTCTTTTAAACCCCTCTTACAAACAGAGCTCCCTT 29083
Oy 3120 ttcagtaagatccgaatcccaatcttaagaacaaagcattagaagaagaatgaatg 3179
Db 29082 TTCAGTAGAAGAGCCGATTCCTCATCTTAAGACAAAGCCATTAGAAAGAGAAAGTAGTG 29023
Oy 3180 agagagagagagaactagctcc 3202
Db 29022 AGAGAGAGAGAGAACTAGCTCC 29000

RESULT 2
AF044216 4818 bp DNA PLN 06-MAR-1998
LOCUS Arabidopsis thaliana steroid 22-alpha-hydroxylase (DMF4) gene,
DEFINITION complete cds.
ACCESSION AF044216
VERSION AF044216.1 GI:2935341
KEYWORDS
SOURCE thale cress.
ORGANISM Arabidopsis thaliana
Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;
Brassicales; Brassicaceae; Arabidopsis.
REFERENCE
1 (bases 1 to 4818)
Choe,S., Dilkes,B.P., Fujioaka,S., Takatsuto,S., Sakurai,A. and
Feldmann,K.A.
The DMF4 gene of Arabidopsis encodes a cytochrome P450 that
mediates multiple 22 alpha hydroxylation steps in brassinosteroid
biosynthesis
Plant Cell 10 (1998) In press
JOURNAL Plant Cell 10 (1998)
REFERENCE Choe,S., Dilkes,B.P., Azpiroz,R. and Feldmann,K.A.
AUTHORS Direct Submission
JOURNAL Submitted (22-JAN-1998) Plant Sciences, University of Arizona,
Tucson, AZ 85721, USA
FEATURES
source
1. 4818
Location/Qualifiers
/organism="Arabidopsis thaliana"
/strain="Ws-2"
/db_xref="taxon:3702"
/chromosome="III"
/map="8 cm from nga6"
990..4398
gene
990..4398
/feature="DMF4"
/note="dwarf4"
/allele="dwarf4-1; dwarf4-2; dwarf4-3; dwarf4-4"
990..1055
TATA_signal
990..1055
/gene="DMF4"
mRNA
join<1133..1353,1434..1758,1844..1996,2095..2409,
2562..2654,2746..2824,2931..3040,3795..4040>
/gene="DMF4"
CDS
join(1133..1353,1434..1758,1844..1996,2095..2409,
2562..2654,2746..2824,2931..3040,3795..4040)
/gene="DMF4"

```

```

/function="brassinosteroid biosynthetic enzyme"
/note="member of the cytochrome P450 superfamily; CYP90B1"
/codon_start=1
/product="steroid 22-alpha-hydroxylase"
/protein_id="AF005093.1"
/db_xref="GI:2935342"
/translating="MEFEHHTLPLPLPLPSLLPLPLILKRRNRKTRNLPKGS
GMPFGETIGYLPYATATLGDPMQOHSKYGIYSNLFGEPTVSADAGLNREILQ
NEGRLEFCSYPRSTGIIIGKMSMLVLDMDHMDRSISLNFSLHRLRTILKVERH
TLFVDSMOONSIFSAODEAKKFPFNIMAKHIMSDDPEEFTEOLKREYVPMKGVY
APLNPSTHAKALDSRATILKFIKKEERKLKEDQEEVEYKTEDEKMSDH
VKRQRTDDLLGWVHLKHSNLSTOILDLILSLFAGHETSVALALAEFLQACPVA
BELREHLEISARKKELGSELMMDYDKMDETCOVINETLGNVVFRLHAKADY
RKGYDIPSGKVLPLVISAVIDNDRYDQPNLFNPMQOONNGASSGSFSYGN
NMPFGGPRILCAGSELAKLEVAVFIHLVYLFKNMELADDQPPAFPFVDFNGLPIR
VSRIL"
BASE COUNT 1566 a 727 c 888 g 1637 t
ORIGIN
Query Match 35.4%; Score 1132; DB 12; Length 4818;
Best Local Similarity 100.0%; Pred. No. 2.1e-137;
Matches 1132; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Oy 2071 gaattcaaatatcccatataatttttgaatctcaaatatataattagcaat 2130
Db 1 GAATTCAAATATCTCATTAATATTTTTCATATCTACAAATTAATATTAATAGCAAT 60
Oy 2131 aacaaatgcatagaagttccaaaaaaatttgttaacagaactcccaattttttt 2190
Db 61 AACAAATGCATAGAAAGTCCAAAAAAATTTGTAAACAGAACTTCCAAATTTT 120
Oy 2191 ttatagaaacagaataacagatagaacataatttgtgttgatggaatggaatga 2250
Db 121 TTTATGGAACAAAGAAATACAGATAGAAAACTATTTTGTGTGGAATGGAATGA 180
Oy 2251 tacattaaacaaatttcaaaaaatlatataagcctacagcgtcaagatgttctca 2310
Db 181 TACATTAAAGCAATTTTAAAAATTTATATAGCCTTATACCGCTCAAGATATTTATCTA 240
Oy 2311 gtatgttaattataatgcatgtgtgcattcagaattggagacaacatgaacagaa 2370
Db 241 GTAGGTGATATTAATATGATGATGATGATGATGATGATGATGATGATGATGAT 300
Oy 2371 taataataacttaataataataataatgtgaatgaatgtgtttctgacattgagg 2430
Db 301 TAAATATTAACTTTTAAATAATAATAATAATAATAATAATAATAATAATAATA 360
Oy 2431 ggcataaaacagaacatgcccagaatcaggttggaactgacgttcgtaataatc 2490
Db 361 GGCATAAAAGACAAATGCCAAAGCTTACGGGTTTGACGTGCAGTCCGTAAATATCT 420
Oy 2491 aataactgtcttgcacgcagcgtcgtgtgtaggggtcctctgacaatttcaactgt 2550
Db 421 AATAACTGTCTTGTGACCGCACGCTGTGTAGGGGTCCTTGACATTTTCACTGTCT 480
Oy 2551 acccctactcgtgagcccaaccccttcccatatcctaaaggttaatttggaaatccca 2610
Db 481 ACCCCTACTCTGTGAGCCACACCTTTTCCATATCTTAAGGTAATTTTGGAAATCCCA 540
Oy 2611 ttaacgcattgagacgcgtacgcgactcctgtgattctctctgagacattacaaaat 2670
Db 541 TTTAACCAGATTGAGACCGTACCGGACTCTCTGGGATTTCTCTGGACATTTATCAAAA 600
Oy 2671 tattagcaagaatgggttatttaattaaacacaaactgacatgaagataaattcat 2730
Db 601 TATTACACCAATAGGTTTATTATTTAAACCTCAACACTGATCAGATGAATTTTCAT 660
Oy 2731 aaacacttttagatgattcgtacgactatcatatgacttttttttttttttttttt 2790
Db 661 AAACACTTTTACATGATGATTCGATCAGATCTATCTAATGACTTTTTCCTACACGG 720
Oy 2791 ggaatgaagttagactatagcagagagaacattgattatagataatccatcaatcc 2850

```

```

|||||
Db 721 GGATGGAAGTTATAGTACTATTAGCCAGACAAATGATTATGATATATCCATTATCC 780
OY 2851 atgataattatgataaataagctgttaaacatttcagatcgagctttcgaact 2910
Db 781 ATGATATTTATGATATTAATAGCTGTGTAACATATTCAGATCGCAGCTTTCGCACTT 840
OY 2911 ttgttttaattaaagatttaataaataaagatttaaaagagataaagagcaaca 2970
Db 841 TTGTTTTTATTAAGAGTTTATTAATAAAGTATTTAAAGAGATATAAGGAGCAACA 900
OY 2971 aaagtaataaacaagagaaacaacaagccatgaagctcatttgtagttaagcttaata 3030
Db 901 AAGGTATGATGAACGAGGAGAAACAAAGCCATGAGCTCATTTGGTGTGTTAAAGCTTAATA 960
OY 3031 agagatttatttaatttaaatgaatgatataaataatttcttaactcttaaa 3090
Db 961 AGAAGATTTTATTAATTTTATTAAGCATGATATACATTTATTTCTGACTTTTAA 1020
OY 3091 acccctctacaacagaagctcccttttcagtagaagctccgattcccaattcaag 3150
Db 1021 ACCCCTCTTACAACAGAAAGCTCCTTTTTCAGTAGAAGTCCGATTCCCAATCTTAAAG 1080
OY 3151 accaagccattagaagaagaagtgaagtagagagagagaagaactagctcc 3202
Db 1081 ACAAGACCATTAGAAGAGAAAGTGAAGTAGAGAGAGAGAGAAACTAGCTCC 1132

RESULT 3
AC005504 104992 bp. DNA HTG 01-APR-1999
LOCUS Plasmodium falciparum chromosome 12, *** SEQUENCING IN PROGRESS
DEFINITION *** 3 unordered pieces.
AC005504
AC005504.3 GI:4558584
VERSION HTG: HTGS_PHASE1
KEYWORDS malaria parasite P. falciparum.
SOURCE Plasmodium falciparum
ORGANISM Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
REFERENCE 1 (bases 1 to 104992)
AUTHORS Hyman,R.W., Fung,E.L., Qin,F., Tamaki,T., Kurd,I.O.B., Conway,A.B.
and Davis,R.W.
TITLE Plasmodium falciparum 3D7 chromosome 12
JOURNAL Unpublished
AUTHORS 2 (bases 1 to 104992)
Hymen,R.W., Qin,F., Fung,E.L., Conway,A.B. and Davis,R.W.
TITLE Direct Submission
JOURNAL Submitted (21-AUG-1998) Stanford DNA Sequencing and Technology
Center, Stanford University, 855 California Avenue, Palo Alto, CA
94304, USA
COMMENT
On Apr 2, 1999 this sequence version replaced g1:4337172.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 3 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
1
58642: contig of 58642 bp in length
* 58643 58842: gap of unknown length
* 58843 91011: contig of 32169 bp in length
* 91012 91211: gap of unknown length
* 91212 104992: contig of 13781 bp in length.
* Location/Qualifiers
1. 104992
/organism="Plasmodium falciparum"
/db_xref="taxon:5833"
/chromosome="12"

BASE COUNT 44286 a 9326 c 9564 g 4141 t 405 others
ORIGIN

```

```

Query Match 4.2%; Score 133.2; DB 60; Length 104992;
Best Local Similarity 44.7%; Pred. 7.6e-09;
Matches 1041; Conservative 0; Mismatches 1248; Indels 39; Gaps 12;

OY 108 ataattaataataataatcaatttaataataga-agtaattaatcattacatcg 166
Db 72449 ATAGACTTATACAAATATTAATAATTAATCAATTAATTAATTAATTAATTA 72508
OY 167 tatgttataacaataatcattccttggtatataatgaaataatgagcttggaattat 226
Db 72509 ATTTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 72568
OY 227 aataataaaggaataatcgaatccatttggtggaat-tacaagattgaagcttgcgt 285
Db 72569 ATTAATTTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 72628
OY 286 tctttggtatataatgataatgagtaatacaaaagagatgagtgagtgaaacatc 345
Db 72629 AATTATTTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 72688
OY 346 tctgtatgaccccccaaaaaaacaacaaaga-taacaaccccccccgatataag 404
Db 72689 ACACATTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 72748
OY 405 ttttggttcggtataggttatttgcataataatataatgacatcattccttgatc 464
Db 72749 AATTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 72808
OY 465 atgaagatttcttaccataataatcgaatctacatccttgatataataataat 524
Db 72809 ATTAATTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 72868
OY 525 acgagtgtaataatccgtttatcgaatccatccatgattgattccttgycatc 584
Db 72869 AATAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 72928
OY 585 cagcaaatattataacaagaagcttgagaacaaacgaacaaataagaagaagaagctg 644
Db 72929 TATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 72988
OY 645 tgaccataggtatgattgataatattcgaagaagaataagaagatgacaaagaagctg 704
Db 72989 ATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 73048
OY 705 tgaataatcgtcccgcaagcttctcccaatcaatcgaacccatttgagattc 764
Db 73049 ATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 73105
OY 765 tggataatcgttaaaatttgcaataagatcgtgaaataatatttatttgtagctgac 824
Db 73106 ATGAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 73165
OY 825 tcaatatattgtcccgagatttgcaatccttcgtttaagcatatttgctctc 884
Db 73166 AAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 73225
OY 885 ttggttcgttcctcctacatataatcgcggaataatgatacaatgataatcgc 944
Db 73226 AATTAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 73285
OY 945 aaaaacaattgctggagcaatttgcaataaacttttccaacaatcagggacactgga 1004
Db 73286 ATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 73343
OY 1005 ctcgacccctaaataacgatttaccagcgtcctcgttgagattcctgcaataagcata 1064
Db 73344 ATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 73403
OY 1065 aaggagccgttcaagctattatacaagaatcacaacgaatagatgctgaaatccctta 1124
Db 73404 AATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 73463
OY 1125 gaaaatttggaattaccggttcgtatgtaaatatagatttagtggtaacaataatgct 1184

```


[illegible][illegible]

| | | | |
|----|-------|--|-------|
| Oy | 759 | atcttcggaatcttcgaaatcttgagataacgattgctggaataatcttcttcgtag | 818 |
| Db | 85664 | TAATTTAAATATTAATTAAG-----AAGATATTTTAACCTTTA | 85623 |
| Oy | 819 | ctgacatcaatactgctccagtgattgataactcttcgtctaaagacatacttgc | 878 |
| Db | 85622 | TTTAAATTAATATAGTAATTAATTAATTTATGTTATTTATTAATTAATTAATTA | 85564 |
| Oy | 879 | ttctcttctgcttcctccttaactatatactcgcgagatatatgaataacatgta | 938 |
| Db | 85563 | TTTTATTTTATTTATTAATTAATTAATTTTAATAATATATTTATTTTAAATTTAA | 85504 |
| Oy | 939 | taaccaacaacatgctcggagccacttcgaataactttctccacaacatlaecggac | 998 |
| Db | 85503 | ATATTAACCATTAATTAATTAATTAATTAATTAATTAATTTTTTTTAATTAATTAAT | 85444 |
| Oy | 999 | actgcactcgcacccttaaaatacgaatttcacagcgctcactgctgagattactagcaaa | 1058 |
| Db | 85443 | ATATTTCCATTTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT | 85384 |
| Oy | 1059 | agccataaagagccgctcaagctatctacaaagctcaaacctgaatagacttgaat | 1118 |
| Db | 85383 | AAATTAATTAATTAATTAATTTGCTGAATTTAATTAATTAATTAATTAATTAATTAAT | 85324 |
| Oy | 1119 | cccttgaanaatcttgagattccggtctgtaactgtaatatagatttagtggtaacaaa | 1178 |
| Db | 85323 | TATATATTAATTAATTAATTAATTAATTAATTAATTAATTAATTTTATTAATTAATTAAT | 85264 |
| Oy | 1179 | ta-----tgttaatcaattagtggtcaacataactaatctcttcagaaaaaacaaact | 1234 |
| Db | 85263 | AAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT | 85204 |
| Oy | 1235 | taagaagctacataatccataatagggatgcatacccttcacagctacgtactacag | 1294 |
| Db | 85203 | TATATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT | 85144 |
| Oy | 1295 | agactaaagaa-----tagttaatgtagtgcgataaatgaataatccacgcggtgtaact | 1348 |
| Db | 85143 | ATAAATATATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT | 85084 |
| Oy | 1349 | aattctggacgcgtatctgcagtaacgcgcgaataatcatctctcgttctggtcgaacataaa | 1408 |
| Db | 85083 | TATATTTATTTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT | 85024 |
| Oy | 1409 | aacaaaaacaagaaaaaaagaaacgacttctctgagatccatcaatgaatcacaatg | 1468 |
| Db | 85023 | ATTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT | 84964 |
| Oy | 1469 | catagactcttgggttcaagttcgaagctcctcacaagcgttgtaaacatctgcacacta | 1528 |
| Db | 84963 | TTTTATTAATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTAT | 84904 |
| Oy | 1529 | ttaaatctgcttcttcaagcatcttcaacataatattctgtagtggagattaaaga | 1588 |
| Db | 84903 | ATAATATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT | 84844 |
| Oy | 1589 | gcgaactctgtaacattacaatacttatatctagctagctagctatgctatcccaatata | 1648 |
| Db | 84843 | TATATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTAT | 84784 |
| Oy | 1649 | tactcttgagctttaaacataactctggttcttcctcgaagctataaatacttaacatcgag | 1708 |
| Db | 84783 | TATTAATTTATTAATTTATTAATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTAT | 84724 |
| Oy | 1709 | gtaaaaaaagcttctgctatcttccgagctgatatgaagataaacctaaatgacttaact | 1768 |
| Db | 84723 | ATATATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTAT | 84664 |
| Oy | 1769 | tttctgaaaatgtaacccttcttaactaataatgaataatcagctatgcttctgctgcataat | 1828 |
| Db | 84663 | ATTAATTAATTTATTTATTAATTAATTAATTAATTTATTTATTTATTTATTTATTTATTTATTTAT | 84604 |

[illegible]

| | |
|------------|--|
| RESULT | 7 |
| AE001398 | |
| LOCUS | AE001398 14867 bp DNA INV 06-NOV-1998 |
| DEFINITION | Plasmodium falciparum chromosome 2, section 35 of 73 of the complete sequence. |
| ACCESSION | AE001398 AE001362 |
| VERSION | AE001398.1 GI:3845197 |
| KEYWORDS | |
| SOURCE | malaria parasite P. falciparum. |
| ORGANISM | Plasmodium falciparum |
| REFERENCE | Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium. |
| AUTHORS | 1 (bases 1 to 14867) Gardner,M.J., Tettehin,H., Carucci,D.J., Cummings,L.M., Aravind,L., Koonin,E.V., Shallom,S., Mason,T., Yu,K., Fujii,C., Pederson,J., Shen,K., Jing,J., Aston,C., Lai,Z., Schwartz,D.C., Pettea,M., Smithberg,S., Zhou,L., Sutton,C.M., Clayton,R., White,O., Smail,H.O., Fraser,C.M., Hoffman,S.L. et.al. Chromosome 2 sequence of the human malaria parasite Plasmodium falciparum Science 282 (5391), 1126-1132 (1998) |
| JOURNAL | 99021743 |
| MEDLINE | Erratum:[published erratum appears in Science 1998 Dec 4;282(5395):1827]] |
| REMARK | 2 (bases 1 to 14867) |
| REFERENCE | Gardner,M.J. |
| AUTHORS | Direct Submission |
| TITLE | Submitted (02-NOV-1998) The Institute for Genomic Research, 9712 Medical Center Drive, Rockville, MD 20814, USA |
| JOURNAL | Location/Qualifiers |
| FEATURES | 1..14867 |
| source | /organism="Plasmodium falciparum" /db_xref="taxon:5833" /chromosome="2" complement(1570..2424) /gene="PFB0490c" complement(1570..2424) /gene="PFB0490c" note="predicted by GlimmerM" /codon_start=1 /product="hypothetical protein" |
| gene | |
| CDS | |

gene
CDS

```
/protein-ig="AAC71887.1"  
/db_xref="gi:3845198"  
/translation="MKENKEKIMDYLSCPLDDVVDREKKSGNSLLKSSSTKKSDYK  
SIFPKSRDSDHKSSGPPGRSGFINRKSQFPPYNNRLINNNYNNGRPHNHR  
DNFKRQSPGSGRVDNMGKSSKRRFTSNRRKSSVSKYRQSGSKMKGRKFNNAAPS  
TYVRLNNYKTVASDPYKCFNNLLSLTKRRRTALTKSKPVGITKSSVPRKRLEK  
GLKGSLSLKSIRKSTSGSKFKPLNKFSLIKITVSLNKLPSPLKEQNIENVNPESL  
NNATIKKN  
11241..14606  
/gene="FEB0495w"  
11241..14606
```

| BASE COUNT | 6284 | a | 1019 | c | 1106 | g | 6458 | t |
|------------|------|---|------|---|------|---|------|---|
| ORIGIN | | | | | | | | |

| | | | | |
|-----------------------|--------------|------------------|------------------|--------------------|
| Query Match | 4.0%; | Score 129.4; | DB 4; | Length 14867; |
| Best Local Similarity | 43.8%; | Pred. No. 3e-08; | | |
| Matches 1038; | Conservative | 0; | Mismatches 1306; | Indels 25; Gaps 10 |

| | | | |
|----|------|--|------|
| Db | 7152 | ATATAAATTATATGCTAAATATATATAATATTAATTAATAATATTAATTAATTAATTAATTAAT | 7211 |
| Qy | 166 | gtatgttaacaacaatactactcttggatagatgagaataatgagcttggaaatta | 225 |
| Db | 7212 | AATATGATATAAATAATATTAATATTAATTAATTAATTTTATATATATTAATTAATTAATTA | 7271 |
| Qy | 226 | laataaagaagaataatcagatccattggttgatcacagctgaagtcttgctg | 285 |
| Db | 7272 | TTATATCTTAATTAATAAATTAATATATTAATTAATTTTATATATTAATTAATTAATTAATA | 7331 |
| Qy | 286 | tcttctgtatagtatatgagtaaatcaaaaagatltgttgaagtgtaacatct | 345 |
| Db | 7332 | TAAACAATATATAAATTAATTAATGCTAAATTAATTAATTAATTAATTAATTAATAATA | 7391 |
| Qy | 346 | tcgtatagaccccccaaaaaaaaaaaaaaacaacaacaaccccccccgatatagt | 405 |
| Db | 7392 | ATATAATATTTAATATATAAACAATAATTAATTTAATTAATTAATTAATTAATTAATA | 7451 |
| Qy | 406 | tttggcttcggaattggttatcttgatcataataatacagctcaattccttgatcta | 465 |
| Db | 7452 | TATTAATAATTAATTAATTAATTAATTAATAAATAATTAATTAATTAATTAATTAATA | 7511 |
| Qy | 466 | tgaagatctctctcttaccaatlaaagtccgaatccatccctctgattataaatt | 520 |
| Db | 7512 | CACGTATATTTTATATATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT | 7571 |

| | | | |
|----|------|---|------|
| OY | 521 | aaahagagiblgaaatccggttatcgataccatccatcaatgatgatgactctg | 580 |
| OY | 521 | | |
| Db | 7572 | TAAATAAAAAAGAAATATATATGTTTATTTAATTTAAATAAATATATATATTTCA | 7631 |
| OY | 581 | aatccagcaaatcttaacaagagatctggaaaaaacgaaataagaagaagggaaga | 640 |
| Db | 7632 | TATATTAATTTAACTTAAATTAATTAATTAATTAATTAATTTCAATTTATACCATTTTAATTT | 7691 |
| OY | 641 | gtagtgaccaatggagatgtagtaatacttaacaagagataagagatgacaaccaag | 700 |
| Db | 7692 | AATATATATAAAAAATATATATTTTAA - ATTAATAATTAATTTAAATTAATTAATTAATTA | 7750 |
| OY | 701 | gtctggaataatgctcccgcaagctcttcacaaatcaatcagaccattlgat | 760 |
| Db | 7751 | TATTTTATATATATTTAATTTAAATATATATATTTAATTTTAAATAAATAATTAAT | 7810 |
| OY | 761 | cttcctggatctgttaaaaattgacgaataacgtctggaataatacttatctgtaagc | 820 |
| Db | 7811 | ATAAATATATTAATAATATTTAATTTAATTAATTAATTAATTAATTAATTAATTAATTA | 7870 |
| OY | 821 | gactcaaatat---tatgctccaggaattctgcaaatctctctgttbaagcaattctg | 877 |
| Db | 7871 | TATATTTAAAAATATAAAAAGTAAATATAAATTAATTAATTTCTTTTAAATTTATTTAT | 7930 |
| OY | 878 | cttcctctctgcttgcgtctctcttaactatatactcgcgagatataagataaagat | 937 |
| Db | 7931 | AATTAATTTATTTTATTTAATTTATTTATATATATATATATATTTATTTTAAATGTT | 7990 |
| OY | 938 | atat-cacaaaaaacaattgctcggaccattctggaataaactcttcacaaatcaggg | 996 |
| Db | 7991 | AAATAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA | 8050 |
| OY | 997 | acctggaatcgagacccttaaaaataagattctaacgcgtaactagttgagatctacgat | 1056 |
| Db | 8051 | ATAATATATTTTACATATAAAAATATATATATATATATATTTAATTTATTTATTTATTT | 8110 |
| OY | 1057 | aaagcataaaggaccgctcaagctcttaacaaagttacaacgtaactgatactagctga | 1116 |
| Db | 8111 | AAATGAATTAATTAATTTATTTATTTATTTAATTTAA---TTTATATTAATTAATTTATTTA | 8167 |
| OY | 1117 | atcccttagaaaaatttgaataaccggtgtgtagtaataagaatagtagtgaagaaca | 1176 |
| Db | 8168 | TATATTTAAATTAATTAATTAATTAATTTATTTAATTAATTAATTAATTAATTAATTT | 8227 |
| OY | 1177 | aatatgttaatcaatctagtggtcaacatatatactatcttcacgaaaaacaactta | 1236 |
| Db | 8228 | ATAAAAAAAATATATTAATTAATTAATGAATATATATTAATTAATTAATTTATTAATTA | 8287 |
| OY | 1237 | agagaatctaacatcatcatatagtgtagtcatatcccttcacgtacgtacactagag | 1296 |
| Db | 8288 | ATTAATAATTTAAAAATATATTAATTTATTTTATTAATTAATTTATATATATTTATTAATG | 8347 |
| OY | 1297 | actaaagaatagtagtagtgcataaaygaatctcacgcgcggtggaabaatlaatg | 1356 |
| Db | 8348 | TCTATTTATTT - TTATTTTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA | 8405 |
| OY | 1357 | gaccgtagtgaacgtacacgcgcaataatactcttggtgtgtcacaataaaacaataa | 1416 |
| Db | 8406 | TTTGTTTCTTGATATTAATATATATTTATTTATTTATTAATTAATTAATTAATTAATTA | 8465 |
| OY | 1417 | caagaajaaagaagaacgatctctctggaattccatcaatgaatcaaaatgcatagctc | 1476 |
| Db | 8466 | TTAATTAATAAATTTATATATATTTATTTCAATTTATTTTAAATATATTTATTAATTAATTA | 8535 |
| OY | 1477 | ctctgggtctacggtcttcgaagctctctacaagcggtcaaccactcgcacaactaaatg | 1536 |
| Db | 8526 | TTTCCATATTAATAATATATATATTTATTTAAAGTTAATATATATTTATTAAT - TTATTTT | 8584 |
| OY | 1537 | ctctctcttaatgcatcttaacatalttaattgtagtggaaattcaaaagagcgaaact | 1596 |
| Db | 8585 | ATAAATAAATAATTTGCTTAATCATTTAATTTAATTAATTAATTAATTAATTTTAAATA | 8644 |
| OY | 1597 | gtacacatcaaatatttatataatgataactagtagtgaattatccaaatcaactctg | 1656 |

/product="Hypothetical protein, PFC0580c"
/protein_id="CAB38968.1"
/db_xref="GI:4493932"
/db_xref="SPTREMBL:O97257"
/translation="MLGLRKRVYLLVSVSLFAYFLKRRHKDNNVETLLINNDIE
KIKIIRHNKCSYIPLPLFIYDSYLYKNIILMWLFKFRKRDKEEYVITNVYRK
KREARLKYNEISQNLFPNKEYIYEIVLESLKYGILSPHLSLITLISIKNISEHVNP
SLYYVGLKLNKHLINERKLYKPKOINSEHTOQAPRNHHNNNNKRPDIINTHSC
KTNISYSYSTYNNMEKNINIDYKVIINHFTEKISYSTYDENCQHTLNNYILNLOTY
DNICRCLTNTNIVYINFYMINILKYCIKNEETILLNTHIDEMKRTINQKNNTWS
LERYIYSEFFKFEENHIYDFEDQNMHILMKENKPFYNSNENHNNIYKXISDNY
FYDHISSSNCSGFKMLKKQOEDNRHILMGKEYPKMSDHEKNNNTGQNIIEK
FDKDLIKIYFLKGLKDDIOILNELNVIYKRLIFECSEKLSIKTNHILEKME
POKDMITILYNSADYNNLRNLTIRKFSNEEENINSPASPELLSKNITVEEDIG
RSNDNTIYNNVDEKRTKTNTNNNDNDNDCISNNDHCSNNDHCSNNDHCSN
NNNNICSNN
YELAKITANNIDYLGKKEKHEKNDENTNLKYKKEPFMIPLANKINRDKRRI
NIYNNVGDIKNNNIIFKRRKNYLFESSISFYKIDYEYKRLIYIDNLIKFGR
YFLKMEHINRKLKRAFPYIYNDEQIINNYHIIHKRHHIHLKQCKKE
IDVYKREKQIINNININIFSYIHKVHNECVYRIEPAKYNKTLLEYVYNLEK
YKRIKKYLTLYLNNNISNDLYPINISYKTIYONKTKRKNRSHITYSKANDIHILF
LTKRIQNCODIESYLYKRYNENKPNLYELVYELHYIYVSEKRRKKFFSTISSP
YSDVNVNSESFTSEFFSSLYLLFLFYHHPDMASXIFKTLVSGLPYVYSLYNNI
MVVCGPKT"
join(7669, .7717,7800, .7829,7912, .7940,8064, .8106,
8302, .8368,8568, .8641,8713, .8812,8924, .9006,9122, .9258,
9369, .9505,9613, .9838)
/gene="PFC0581w"
join(7669, .7717,7800, .7829,7912, .7940,8064, .8106,
8302, .8368,8568, .8641,8713, .8812,8924, .9006,9122, .9258,
9369, .9505,9613, .9838)
/gene="PFC0581w"
/note="PFC0581w, hypothetical protein, len: 324 aa,
revised: new gene prediction, splicing prediction very
tentative"
/codon_start=1
/product="hypothetical protein, PFC0581w"
/protein_id="CAB90285.1"
/db_xref="GI:7711066"
/translation="MKRRKRVFSIKFVNVKNNVYLSKSDYIRIINMIGILFRL
/IRAFIHYENQJASFSFESHKNVRYTKEDIIDIGIKCFENITIDYISIKQDNFS
NIKTIVTAMKNMQLTTSYSNKKIDIVNASFQWASPEYFLIKRFSHWSPPGA
LWIKDEIYAKNNKMFSSALSDNSNSVTKXIYDLDLNNIIESFKNRPAASGRVYA
TLKEKKRKTINRLLSLSEKYPNNQWYMDMKERKIHITITVTINLFPLLISLRYHDSV
QNFLEKKNNSDLQDDIDDEDEKTYDEEILREAKRSEYDDEEL"
7790, .7799
/note="potential splice acceptor sequence for exon 2 of
PFC0581w"
7830, 7835
/note="potential splice donor sequence for exon 2 of
PFC0581w"
7901, .7911
/note="potential splice acceptor sequence for exon 3 of
PFC0581w"
7941, .7947
/note="potential splice donor sequence for exon 3 of
PFC0581w"
8055, .8063
/note="potential splice acceptor sequence for exon 4 of
PFC0581w"
8107, .8112
/note="potential splice donor sequence for exon 4 of
PFC0581w"
8293, .8301
/note="potential splice acceptor sequence for exon 5 of
PFC0581w"
8369, .8374
/note="potential splice donor sequence for exon 5 of
PFC0581w"
8559, .8567
/note="potential splice acceptor sequence for exon 6 of
PFC0581w"
8642, .8647

[illegible]

ORGANISM Plasmodium falciparum
Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
REFERENCE 1 (bases 1 to 86827)
AUTHORS Bowman, S., Lawson, D., Basham, D., Brown, D., Chillingworth, T., Churcher, S., Craig, A., Davies, R. M., Devlin, K., Felwell, T., Genesby, S., Gilliam, R., Hamlin, N., Harris, D., Holtroyd, S., Hornsby, T., Horrocks, P., Jagsels, K., Jassal, B., Kyes, S., McLean, J., Mould, S., Muller, K., Murphy, L., Oliver, K., Quail, M. A., Rajandream, M. A., Rutter, S., Skellon, J., Squares, R., Squares, S., Sulston, J. E., Whitehead, S., Woodward, J. R., Newbold, C. and Barrell, B. G.
TITLE The complete nucleotide sequence of chromosome 3 of Plasmodium falciparum
JOURNAL Nature 400 (6744), 532-538 (1999)
MEDLINE 99376085
REFERENCE 2 (bases 1 to 86827)
AUTHORS Bowman, S., Skellon, J., Churcher, C., Lawson, D., Quail, M. and Barrell, B.
JOURNAL Unpublished
REFERENCE 3 (bases 1 to 86827)
AUTHORS Lawson, D., Bowman, S. and Barrell, B.
TITLE Direct Submission
JOURNAL Submitted (17-DEC-1998) P. falciparum Genome Sequencing Consortium, The Sanger Centre, Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SA, UK
COMMENT On May 5, 2000 this sequence version replaced gi:4493931. For more information about this sequence or the Malaria Project, see http://www.sanger.ac.uk/Projects/P_falciparum.
FEATURES
source location/Qualifiers
1. 86827
/organism="Plasmodium falciparum"
/strain="3D7"
/db_xref="taxon:5833"
/chromosome="3"
/clone="MAL3P5"
/gene="PFC0575w", MAL3P5.1"
/join(322..603,826..1023,1197..1301,1458..2942)
/gene="PFC0575w", MAL3P5.1"
/note="PFC0575w (MAL3P5.1), Hypothetical protein, len: 689 aa, possible signal sequence, revised: added new exon 2"
/codon_start=1
/product="Hypothetical protein, PFC0575w"
/protein_id="CAB38969.2"
/db_xref="GI:7711065"
/db_xref="SPTREMBL:O97258"
/translation="MYLKNVYIYSSCFILPDLCSFHLKMKYKNHNNKSVTFELRSPOYRKRFRSRKKNVSEFKKOKKPLFLENLKGSFLGFWKNDYDKYIDVISNINNLTRIKOYTHKRSKSEFTKENKQILLCVSKIDFKIINNLSYIKRPOMSNITVHSILNOISEKKEKDAENYLAHLFLKIDENITLFPMMHIMDFEFSKQVCIIRDISKRRKKNLSIYNLFICILYETFCMLILYISHLCEFFFEFFCFGLYNILERTYFECVGDILIRKTERVNLVCEKKIKRPMKDAIKKMKNDKIDYFNHYDELICRFTMKLNTFRNNKNIIRSYNDINNDISIDKDMANRPIDVAINNINISLDEKJKEQFENDENKELKDYEQPLFNNDIKIYIEDOPLVYNDSNNNDNNNNINTKNNKRIKDYVNDDDDYDEKEDLVIOKNIDYIYKNTIGNKSLSLEKKNQFIEQADIEFONFLSNVNDQGRVKSNDENTKSTEHKIKNKTINNGYETLIONMENNFIKKNIDNINSDISNNNEINIRKLKLNOSDEDINTSLDYERLPRKVLWYIOKIEYLRKRYOYDIINEOYPIIKNEKTVLDLNTGYKIVMSPDVNSLPEFKIKIDISIPREKKNQNMENQKSKNYIIMENVM"
misc_feature 604..609
/note="predicted splice donor sequence for exon 1 of PFC0575w (revised)"
misc_feature 816..825
/note="predicted splice acceptor sequence for exon 2 (revised of PFC0575w)"
misc_feature 1024..1029
/note="predicted splice donor sequence for exon 2 of PFC0575w (revised)"
gene /gene="PFC0580c", MAL3P5.2"
complement(3354..6644)
/gene="PFC0580c", MAL3P5.2"
/note="PFC0580c (MAL3P5.2), Hypothetical protein, len:

1097 aa, possible signal sequence, predicted using hexon.
/codon_start=1
/product="Hypothetical protein, PFC0580c"
/protein_id="CAB38968.1"
/db_xref="GI:4493932"
/db_xref="SPTREMBL:O97257"
/translation="MLGLKRNPFYLLVSPSLFAYFLKRRHNNENYETLINNDIEKIKRIRNKCSYIPLFLINLYDSYIKNLIRMLYEFERRRKRQKEYYIYTNVVRKRREAIKRYNLESDQNLFNKRYIETIYELKAGIISLPHILKILISHCVAIYSLIYNKLDKNNHNLNEKKLKYFPOINNEHQOAPNTNTHNNNNNNKKPDIINSHCKNTNISSTYTNMNEKENINIDYKNINHEFTESIYKDNCOHITLMYLLNQTYNDRICLNTNTINIIYINFMINILKYICQKMEIILLYNNHIEDKKTINQKNNTSLEKYIYSEFFKKEENHIYDFEDQMHLKKEKDYKYNSENETHNNIYKISDNYFDHINSSNRCSPFKNLKQOQDQDTKHIIKGKEYPNKKSDHEKKNNGYINIEKDOKKDILKIIYELFKNGKIDDIOLINELIYVMYRLPFCSLKISIKNNHLLKEMEFSDKDKITILNSADTMNLRNRIILKRSKREBERINSPASFPLISKNITIFDEBICRSRQNTIYNVYDEKTNKTTNNNNNDNDNDCSNNDHICSNNDHICSNNDHICSNNNICSNNNNICSNNNNICSNNNNICSNMKMLDEFCQDNKFNDRTRKREKRIYELAKIYNNIYDFYLGKREKHONEDNTINLYIKKCPMAYFLYLNKINDTSEIHNNNIYNGDINKNNIIFKKYMLPESIIISYEKIDYENKRLRYIYDMLIKKFCRYELKMEHINRKLKMKRAPHYIYNFPOFIINNYIHHKNNIHKIHHKQCKDEIDIKFKDLYICMLNININISYIHKVDHNECVIRIFAYKILILEYNTLNENKNTIYKNIKIKYTLNANNISMDLYPNISYKRIYNQNKYKRNKFSHIFYSLKNDIHLFLYTQRIQNCDFISYIYKKNYNEKNPFLNLYELAHYVSEKKKKKFFSSTSSPYSDTMVNSFTFSEFFSLYLLFILFHPDMYASYIFFKTLTYSGLPTYYSYNNIMVCGPKR"
gene /join(7669..7717,7800..7829,7912..7940,8064..8106,8302..8368,8568..8641,8713..8812,8924..9006,9122..9258,9369..9505,9613..9838)
/gene="PFC0581w"
/join(7669..7717,7800..7829,7912..7940,8064..8106,8302..8368,8568..8641,8713..8812,8924..9006,9122..9258,9369..9505,9613..9838)
/gene="PFC0581w"
/note="PFC0581w, hypothetical protein, len: 324 aa, revised: new gene prediction, splicing prediction very tentative"
/codon_start=1
/product="Hypothetical protein, PFC0581w"
/protein_id="CAB90285.1"
/db_xref="GI:7711066"
/translation="MKKKKHYSIKFVNDKNNKYLCSKDYIRIINMIGLHIFRLOKIHIYPANQIASSPFESHKNRYVEDIIDGKEKFNITDILIESIKQDNDFSNDIKTTVAMKNNKMDOLITSYNNKIDITVANSSOMQASPYIFLNIFKFSWSSPGA LKVDDEIVSKKNNPSALSDNSGYKXIYVTLTDNIIEPTKRNPSGVCKVYTLKREKTIHRLISKEKIPNOMQWMDMKRKHITTFVTINLPELISLSHRHDSVQNFLEKEKNNSDKLQDDIDEDERYFDEILIREAKKSEYDKODEL"
misc_feature 7790..7799
/note="potential splice acceptor sequence for exon 2 of PFC0581w"
misc_feature 7830..7835
/note="potential splice donor sequence for exon 2 of PFC0581w"
misc_feature 7901..7911
/note="potential splice acceptor sequence for exon 3 of PFC0581w"
misc_feature 7941..7947
/note="potential splice donor sequence for exon 3 of PFC0581w"
misc_feature 8055..8063
/note="potential splice acceptor sequence for exon 4 of PFC0581w"
misc_feature 8107..8112
/note="potential splice donor sequence for exon 4 of PFC0581w"
misc_feature 8293..8301
/note="potential splice acceptor sequence for exon 4 of PFC0581w"
misc_feature /note="potential splice acceptor sequence for exon 5 of PFC0581w"
misc_feature 8369..8374
/note="potential splice donor sequence for exon 5 of PFC0581w"
misc_feature 8539..8567

| | | | |
|----|-------|---|-------|
| Oy | 1544 | taigcactttaaacaattatcttcttgtagaatttaataagcgagacttgaacct | 1603 |
| Db | 39930 | TAATTATATATGCAATTTAAATGCTATTTA-TAATATTATTAATTAATTAATTTATTTT | 39872 |
| Oy | 1604 | tacaaatttatacttagatactagatctgcatctccaaatacactcttgatgttta | 1663 |
| Db | 39871 | TAATTATATTTTAAATATTTATTTAATTTAATTTAATTTAATATATTAAGTGA | 39812 |
| Oy | 1664 | aacttaactctgtctctccctcaggtataaatacattcaactcgaggtaaanaagtttg | 1723 |
| Db | 39811 | ATMAAATATAAATATATTTATTT-----TATATATATATATTTATTTATTTATMACT | 39760 |
| Oy | 1724 | tctattctcgagatcgatgaagaataaccatgacttcaattcttttgaaatgtaac | 1783 |
| Db | 39759 | GCTTATTTAATTCATTTTAAATTAATATATATATTAATTTGCAATTTATTAATTTATTT | 39700 |
| Oy | 1784 | cccttactcatagatlaattacgcatgtcttctgtcgcaataagacagcccttacaac | 1843 |
| Db | 39659 | TATMAAATTTATATATTAATMAAACCTTTATTTAT-TTAATATTTAATTAATTAATATA | 39644 |
| Oy | 1844 | tgtagtagcaattttctcgcgaatataaattggaaattcaatgcatcactcaatgta | 1903 |
| Db | 39640 | TATTAATATATATATTTATTGCAATTTAATTAATTTATTTATTTTAAATTAATAT | 39581 |
| Oy | 1904 | agaaacagcctgaglatcacatttcaatttgaagacaaaattcttgaaaaatgtlaaatt | 1963 |
| Db | 39580 | ATATATTTATTTATTTATTTTAAATTTATTTATTTATTTATTTATTAATTAATTAATATA | 39522 |
| Oy | 1964 | tctacaatatatataaataatagatgcccataatgatttccctagcttctaataattt | 2033 |
| Db | 39520 | TTAATTTATTTATTTATTTAATTAATTAATTTATTTATTTAATTAATTAATTAATTAATTT | 39461 |
| Oy | 2024 | tctttatatttagtataaatacactatgaaccaaataagtcgtgtaattcaaatatc | 2083 |
| Db | 39460 | TATTTATTTTAAATTAATTAATTAATTAATTTAATTTATTTTATTTAATTAATTAATATA | 39400 |
| Oy | 2084 | tccatttaatatcttttggaaatccacaattataatattagtcataaacaatgcacata | 2143 |
| Db | 39400 | TTAATTTATTTTATTTTAAATTAATTAATTAATTAATTAATTTATTTTATTTAATTAATTA | 39341 |
| Oy | 2144 | aagttccaaaaaaattctgttaaccaggaactccaaattctttctttagagacaag | 2203 |
| Db | 39340 | TATATATTAATTTATTTTATTTTAAATTAATTAATTAATTAATTTATTTTATTTTAAAT | 39281 |
| Oy | 2204 | aaatc---acagatagaasaaactatctgtctggaatggaagtagataatacaatgaac | 2260 |
| Db | 39280 | AAATTTATATATATTTATTTATTTTATTTTAAATTAATTAATTAATTTATTTATTTT | 39221 |
| Oy | 2261 | aaatttbaaaaattataaagccatacagcgctaaagtagtatactagtagtgtaa | 2320 |
| Db | 39220 | AATTAATATATATTTTATTTTATTTAATTTAATTTAATTTAATTTAATTTATTTGTTAT | 39161 |
| Oy | 2321 | ttaa--taatgcatacgatcgactcgaatctggagacaacaatgaaaacsgaatlaaatac | 2378 |
| Db | 39160 | TTTATTTATTTAATTAATTAATTAATTTATTTATTTATTTATTTATTTATTTATTTAT | 39101 |
| Oy | 2379 | taactttaaataaataaaatttgagcaaatatgctttcttgacat | 2425 |
| Db | 39100 | TTAATATTAATTAATTAATTAATTAATTTTATTTTATTTATTTTATTTTATTTTATTTT | 39054 |

| RESULT 10 | | | | |
|------------|---|-----------|-----|-----|
| LOCUS | DMU11584 | | | |
| DEFINITION | DMU11584 | 4601 bp | DNA | INV |
| ACCESSION | Drosophila melanogaster Oregon-R mitochondrial A+T region. | | | |
| VERSION | U11584 | | | |
| KEYWORDS | U11584.1 | GI:508826 | | |
| SOURCE | mitochondrial DNA; A+T region; tandem repeats. | | | |
| ORGANISM | fruit fly. | | | |
| | Mitochondrion Drosophila melanogaster | | | |
| | Eukaryota; Metazoa; Arthropoda; Insecta; Hexapoda; Insecta; | | | |
| | Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; | | | |

| | |
|-----------|---|
| REFERENCE | Muscomorpha: Ephyridioidea; Drosophilidae: Drosophila. |
| AUTHORS | 1 (bases 1 to 4601) |
| TITLE | Lewis,D.L., Farr,C.L., Farguhar,A.L. and Kaguni,L.S. Sequence, Organization and Evolution of the A+T Region of <i>Drosophila melanogaster</i> Mitochondrial DNA |
| JOURNAL | Mol. Biol. Evol. 11, 523-538 (1994) |
| MEDLINE | 94285822 |
| REFERENCE | 2 (bases 1 to 4601) |
| AUTHORS | Kaguni,L.S. |
| TITLE | Direct Submission |
| JOURNAL | Submitted (28-JUN-1994) Laurie S. Kaguni Ph.D, Dept. of Biochemistry, Michigan State University, East Lansing, MI, 48824-1318, USA |
| FEATURES | Location/Qualifiers |

[illegible]

| | | | |
|----|------|---|------|
| QY | 1311 | atggaagtcgtgaataaagaattcaacagcggtgaataaatt--atggagcgtacgtt | 1367 |
| Db | 3521 | AAATTTATTTAAATTAATTAATTAATTAATTTTAAAAATTTCTTAATGATTAATTTTAT | 3580 |
| QY | 1368 | acgcatacgtcaaatatcatctcttggttggtcaacataaacaacaaagaataaa | 1427 |
| Db | 3581 | AAAAAATATTTATTAATAATAAAACATGTTTTTAAAAATAAACAAAAATTTTAAATA | 3640 |
| QY | 1428 | gaacagattttcttggtatccatccatgatacgaataatgatagatccttgggttac | 1487 |
| Db | 3641 | ATTAATTTTATTAATGAATTAATTAATTTATTTTCAATTTTTTTTAAAAAATTTTTTAA | 3700 |
| QY | 1488 | agttcgaagctcctctcaacgcgtgcacacatcgacataataattgcttct-ctaa | 1546 |
| Db | 3701 | AAAAAATATATTTTTTTTTTAAAAAAACGATATACTATTTATTAATTAATGAATATTTA | 3760 |
| QY | 1547 | tgcattcttaacatatattcttggttggtgaatttaataagcgaaacttgatactac | 1606 |
| Db | 3761 | TATATATATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA | 3820 |
| QY | 1607 | aatacttaactagatactagtgatattcccaatacactacttggtggttcaac | 1666 |
| Db | 3821 | AATTTAAAAATGTGATATATATTTATTAATAAATTTTATATTCATATTTATTTATTTAA | 3880 |
| QY | 1667 | ctaaactgttctctccacgataaataatcaatcatcaggtgaataaaagtttgcct | 1726 |
| Db | 3881 | TTTTATTTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT | 3940 |
| QY | 1727 | tatttcgcgatacgaagaataacctaagtaacttaattcttcttgaaatgtaacct | 1786 |
| Db | 3941 | ATTTATTTGAATTTTATTAATTAATTAATTAAT-ATTAATGAATAAATTAATTAATTAATA | 3999 |
| QY | 1787 | cttaactagatattaccgtaigtcttggtgcctaataagagcctcacaacgt | 1846 |
| Db | 4000 | TTTTATTAATTAATTTTAAAAAATTTCTTAATGCTATTAATTTTATTAATAAATAATTTAT | 4059 |
| QY | 1847 | gatagcatactttctcgcaaatatlaaattaggaattcaatgctactatacagaaga | 1906 |
| Db | 4060 | AATTAATAATCATTTTTTTTAAAAATAACAAAAAATTTTAAATAATAATTTTAAATG | 4119 |
| QY | 1907 | aacgcgtgaglatcacattlaaattlaaagcaaaacttctgaaaatgtlaaattct | 1966 |
| Db | 4120 | AAATATATTAATTAATTTTTCATTTTTTTTAAAAAATTTTAAAAAATAATTTT | 4179 |
| QY | 1967 | aacaaatattaaataatgatacctataatgatttctactagtctcttaaatattt | 2026 |
| Db | 4180 | TTTTAAAAAAACGTATATCTAATTAATAAATTAATTAATTAATTAATTAATTAATTAAT | 4239 |
| QY | 2027 | tttaattattgataaatacatatgaaccaaataatgttggaatcaataatccc | 2086 |
| Db | 4240 | TTAATATATTTTATTAATATCTAATTAATTTAAATAAAAAATTTTAAATTTTAAATTAAG | 4299 |
| QY | 2087 | altaaatatttcttgaaacttacaatattataattattgltcaataacaaatgataaga | 2146 |
| Db | 4300 | ATATATATTAATAAATTTTATATTCATATTAATTAATTAATTAATTAATTAATTAATAA | 4359 |
| QY | 2147 | ttccaaaaaaacttcttcaacgaactccaatttttcttattatgacaagaata | 2206 |
| Db | 4360 | TAAATATATGTTTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA | 4419 |
| QY | 2207 | taacagatagaagaacttcttggttggaatggaatgataatacattgaagaactt | 2266 |
| Db | 4420 | TATATATATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA | 4479 |
| QY | 2267 | taaaaaaatatataagcctatacgcgtcaaaatgltatctaglaagtgtaataa | 2326 |
| Db | 4480 | AAAAAATTTCTTAATGATTAATTTATTAATAAATAATTAATTTTAAAAAATAATTTTAA | 4539 |
| QY | 2327 | atgcattgctgcattcgaattgggacacaatgaacgggaattaaatatiaacttta | 2386 |
| Db | 4540 | TTTTTTAAAAAATAATGATTAATTAATTAATAAATAAATAAATAAATAAATAAATAATTA | 4599 |

```

RESULT 11
LOCUS      DMU37541      19517 bp      DNA      circular      INV      02-MAR-2001
DEFINITION Drosophila melanogaster complete mitochondrial genome.
ACCESSION  U37541
VERSION     03/7541.1  GI:1166529
KEYWORDS
SOURCE
ORGANISM   fruit fly.
            Mitochondrion Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE  1 (bases 12511 to 12682)
            Clary,D.O., Goddard,J.M., Martin,S.C., Faureon,C.M. and
            Wolstenholme,D.R.
            Drosophila mitochondrial DNA: a novel gene order
            Nucleic Acids Res. 10 (21), 6619-6637 (1982)
TITLE      2 (bases 5269 to 5695)
            Clary,D.O., Wahlthner,J.A. and Wolstenholme,D.R.
            Transfer RNA genes in Drosophila mitochondrial DNA: related 5'
            flanking sequences and comparisons to mammalian mitochondrial tRNA
            genes
            Nucleic Acids Res. 11 (8), 2411-2425 (1983)
REFERENCE  3 (bases 404 to 5272)
            de Bruijn,M.H.
            Drosophila melanogaster mitochondrial DNA, a novel organization and
            genetic code
            Nature 304 (5923), 234-241 (1983)
JOURNAL    4 (bases 804 to 1778)
MEDLINE    83245048
AUTHORS    Saita,Y., Ishiwa,H. and Chigusa,S.I.
TITLE      Analysis of nucleotide substitutions of mitochondrial DNAs in
            Drosophila melanogaster and its sibling species
            Mol. Biol. Evol. 4 (6), 638-650 (1987)
JOURNAL    5 (bases 5268 to 13619)
MEDLINE    88174373
AUTHORS    Garesse,R.
TITLE      Drosophila melanogaster mitochondrial DNA: gene organization and
            evolutionary considerations
            Genetics 118 (4), 649-663 (1988)
JOURNAL    6 (bases 441 to 2967)
MEDLINE    88212147
AUTHORS    Saita,Y. and Takahata,N.
TITLE      Evolution of Drosophila mitochondrial DNA and the history of the
            melanogaster subgroup
            Proc. Natl. Acad. Sci. U.S.A. 87 (24), 9558-9562 (1990)
JOURNAL    7 (bases 14215 to 14512)
MEDLINE    91088557
AUTHORS    Ballard,J.W., Olsen,G.J., Faltch,D.P., Odgers,W.A., Rowell,D.M. and
            Atkinson,P.W.
            Evidence from 12S ribosomal RNA sequences that onychophorans are
            modified arthropods
            Science 258 (5086), 1345-1348 (1992)
JOURNAL    8 (bases 14917 to 19517)
MEDLINE    93088057
AUTHORS    Lewis,D.L., Farr,C.L., Farquhar,A.L. and Kaguni,L.S.
TITLE      Sequence, organization, and evolution of the A+T region of
            Drosophila melanogaster mitochondrial DNA
            Mol. Biol. Evol. 11 (3), 523-538 (1994)
JOURNAL    9 (bases 1 to 408; 13319 to 19517)
MEDLINE    94285822
AUTHORS    Lewis,D.L., Farr,C.L. and Kaguni,L.S.
TITLE      Drosophila melanogaster mitochondrial DNA: completion of the
            nucleotide sequence and evolutionary comparisons
            Insect Mol. Biol. 4 (4), 263-278 (1995)
JOURNAL    10 (bases 1 to 19517)
MEDLINE    96423163
AUTHORS    Lewis,D.L., Farr,C.L. and Kaguni,L.S.
TITLE      Direct Submission
            Submitted (03-OCT-1995) Laurie S. Kaguni, Biochemistry Department,
            Michigan State University, East Lansing, MI 48824-1319, USA

```

```

FEATURES
source
location/Qualifiers
1..19517
    /organism="Drosophila melanogaster"
    /organelle="mitochondrion"
    /db_xref="taxon:7227"
    /note="derived from new and previously submitted
    sequences; sequence is a composite containing sequences
    obtained from different Drosophila melanogaster strains"
1..65
    /product="tRNA-Tle"
    /complement(97..165)
    /product="tRNA-Gln"
171..239
    /product="tRNA-Phe"
240..1265
    /codon_start=1
    /transl_table=5
    /product="NADH dehydrogenase subunit 2"
    /protein_id="AAC47811.1"
    /db_xref="GI:1166530"
    /translation="MFNNSKILFTITIGTLITVTSNSWIGAMGILEINLSRPL
    LSDNNNMSTASLKYRTQVLAVILFSSILLMKNNNNNESTSMINSALL
    LKSGAPFHEWEPNNMEGLTWNNALMTWOKIAPMLITLNTIKLILSVILSVII
    GAIGLNOTSLRKLMAFSSINHLGWLMSLSMISBIMLIFEFYSFSEVLTPEMFI
    KLFLNQLFSEFVNSKILKFLFNNFLSIGLPPLGFLPKMLVIOQLTLCNOYEMLT
    IMMSTLITLFEYLRLICYSAPMNNYFENNIMKMNNSINYNMIMTFISIFGLFI
    SLFVFMF"
1264..1329
    /product="tRNA-Tyr"
    /complement(1322..1383)
    /product="tRNA-Cys"
    /complement(1403..1468)
    /product="tRNA-Tyr"
    join(1470..1472,1474..3009)
    /note="mechanism underlying reading frame shift after
    first codon uncertain"
    /codon_start=1
    /transl_table=5
    /product="cytochrome c oxidase subunit I"
    /protein_id="AAC47812.2"
    /db_xref="GI:7412849"
    /translation="MSROWLSTNHNKDIGLTFIFGAGAVGTSLSILINAEIGHPC
    ALIGDDQIYNIVTAHAETIMIFPVMPIMIGFEGNWLVPMLGADNAPPRNNNSFW
    LLPPALSLLLVSNVENGAGCTGWTVPPLISAGIAGGASVDLAIFSLHAGISSILGA
    VNEITVYINRSTGSLSDRMLFVWSVITALLLSLPVLAGATMLLTDRNMTSE
    FDPGGGDPILLYOHLFMPFGHEVYILILPGGMSHIIISQSGKKEFGSLGMYAM
    LAIILGRTVAHHFTYGMVDVDRAYTSMTIAYPTGIKITSMTLTLCTOISYS
    PAIIMALGEVFLFTVGLTGVLANSSVDIILHDITVVAHFHYLSGAVFALMAGF
    IHWPPLFTGLTLNKKMLKSHPTIMEIGVNLFFPDHFFQHGMPRRSYDPAAYTWN
    IVSTIGSTISLIGILFEFFIIMESLVSQROYIYPIQLNNSIEMVQMPASHSELP
    LLTN"
3012..3077
    /product="tRNA-Leu"
3083..3767
    /note="TAA stop codon is completed by the addition of 3' A
    residues to the mRNA"
    /codon_start=1
    /transl_except=(pos:3767,aa:TERM)
    /transl_table=5
    /product="cytochrome c oxidase subunit II"
    /protein_id="AAC47813.1"
    /db_xref="GI:1166532"
    /translation="MSTWANGLODSASPLMEOLIFFPDHALLLIVMTIVAGLMMF
    LFPNNVNRFLHGOILIEWTILPAIILFIALPSRLALYLDEINPSVTLISIGH
    OMWYSYSDPNNIEFSDYMTPTNELMDGRLLDQDNRVYLPKNSQIRILVTADYDI
    HSWTVPALGAVVDSTPGRLNOTNEFINRPLFYGQCSBICGANSFMPIVIESPVN
    FIKWISSNNS"
3768..3838
    /product="tRNA-Lys"
3840..3906
    /product="tRNA-Asp"
3907..4068
    /codon_start=1

```

```

/transl_table=5
/product="ATPase 8"
/protein_id="AAC47814.1"
/db_xref="GI:1166533"
/translation="MPQMAPISWLLIFIIISITPLFCISINYISYMPSPKSNELKNI
NLNSMNMW"
/codon_start=1
/4062..4736
/transl_table=5
/product="ATPase 6"
/protein_id="AAC47815.1"
/db_xref="GI:1166534"
/translation="MMTNLFVDFPLAIFNFSNLWLSFGLMLIPSYIWMPSRYNI
MNSILTLHKEFKTLPGSCHNGSTFIIFLSLILNNMGLEPFIYFTSHTLFL
LSLALPLMELGMLGWIMNHQMFALHVPQGPATIMFNCIENTISIMPTGTAVR
LTPANINAGHLILLTLLGNTGSSMYMLMFLMAQIALLVLESAAVMIQSYFAVSTL
YSSEYN"
/4736..5524
/codon_start=1
/transl_table=5
/product="cytochrome c oxidase subunit III"
/protein_id="AAC47816.1"
/db_xref="GI:1166535"
/translation="MSTHSNHPFLVDYSPMPPLTGALGAMTTVSGVWVWFOXDLSL
VLGNITLIVYQWMDVSRREGTVOGLHTVATVIGLRMGMLIFLSEVLFVSPFMAF
FHSLSPAIELGASWPMGIISFNFOIPLNTAILLAGVTLVAHSLMLNNSOT
TOGLFTVLTIGVETIIQAYEYIEAPFTIADISYVSTPEMAGFHIVLIGTFLV
CLRLHNHFSKNHFGFEAAMWVHFVYVWVLFITITIMWG"
/5543..5607
/product="cRNA-Gly"
/5608..5961
/codon_start=1
/transl_table=5
/product="NADH dehydrogenase subunit 3"
/protein_id="AAC47817.1"
/db_xref="GI:1166536"
/translation="MESIFIALILLITTYIMFASILSKALIDREKSPPECQFD
PKSSRLPELRFPLITIIPLIFDEIALILPMIIMKYSNMIMWTISILFILLI
GLYHMGNMNSN"
/5981..6045
/product="cRNA-Ala"
/6055..6118
/product="cRNA-Arg"
/6119..6183
/product="cRNA-Asn"
/6184..6251
/product="cRNA-Ser"
/6252..6318
/product="cRNA-Glu"
/complement(6337..6401)
/product="cRNA-Phe"
/complement(6401..8124)
/note="TAA stop codon is completed by the addition of 3' A
residue to the mRNA"
/codon_start=1
/transl_except=(pos:complement(6401..6402),aa:TERM)

Query Match      3.9%: Score 124.4, DB 6; Length 19517;
Best Local Similarity 43.9%: Pred. No. 1.3e-07;
Matches 1027; Conservative 0; Mismatches 1276; Indels 37; Gaps 10;

```

```

246 gattccattggtggtacacagcttaagtttctgcttcttctgtatatactatcg 305
17372 TTTATTACCAATAATTAAATTAATAAATAAATTTATTTTATTTTATTTTAA 17431
306 agttaaatacaaaagatctgattgaagtgtaacatattcgatagcccccacaaa 365
17432 AATTAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 17491
366 aaaaaaaacaaacaaaccccccccgatataagtttctgcttctgattagctt 425
17492 ATATATATGAAAAATTA-----AATTTTAAATTAATTAATTAATTAATTA 17536
426 tatttcataataatcacatgcatcttcttctgattcattcaagaatttcttccaa 485
17537 TTTTAAAAATTTCTTAATGCTATTTATTTTAAATAAATTTATTAATTAATTA 17596
486 aaaaattcgaatcatatctctgatttaataataaagaagtgtaataatcgctta 545
17597 TTTTAAAAAATAACAAAAAATTTTAAATAAATTTTAAATAATTAATTAATTA 17656
546 tggatcaccccaatcatgattatgattcttgctatccagcaattatcaagaagt 605
17657 TATTTTTCATTTTAAAAAATTTTAAAAAATTAATTTTAAAAAATAA 17716
606 attgaagaaaaacccgaataaagaaggaagtaagtgaccatcggagtaagta 665
17717 CATATACATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 17776
666 aattalcaagaagaataagaatgacaacaaaaggctgtaagaaatggtccctgc 723
17777 ATATATCTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 17836
724 agcttctctcaacaatcatatcgacctatctggttcttctggaatcgttaaat 783
17837 AATTTATATTCATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 17896
784 ggcataagattgtaaa--aaatttatttgtagcgaatcgaatcaattatgt- tcca 840
17897 TAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 17956
841 ggtattgataatctctctgttaaaagcatatttcttcttcttctgcttctct 900
17957 TATATATGAAAAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 18016
901 taactataatcatcgcgatata-----gataacaatgatatcacaaaatgt 955
18017 AATGTATTAATTTTAAATAAATAATTAATTAATTAATTAATTAATTAATTA 18076
956 ctggagccatttgaataacttcttccaacattcggagacacggaatcgaccct 1015
18077 AAAAAATTTTAAATAATTAATTAATTAATTAATTAATTAATTAATTAATTTT 18136
1016 aaatcagatttcaagcgtcatctgattgatactacgataagaataagaccgt 1075
18137 TTTAAAAAATTTTAAAAAATAATTTTAAAAAATAATTAATTAATTAATTAATTA 18196
18137 TTTAAAAAATTTTAAAAAATAATTTTAAAAAATAATTAATTAATTAATTAATTA 18196
1076 ca-----agctatttacaagaagttacaacgtgaatcgtgaatcccttaaga 1130
18197 AATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 18256
1131 ttggaattacccggtctgtatgtaaatatagatttggttaacaataatgtaacaa 1190
18257 AATTAATAAATTTTAAATAATTAATTAATTAATTAATTAATTAATTAATTTCT 18316
1191 ttagtgcaacatataatcttcttcaagaataaacaacttaagaagttaaat 1250
18317 AATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 18376
1251 atccatatatgggtatgctatcttcaagttatgctatcagtaagaataagta 1310
18377 AATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 18436
1311 atgtgatgctgataaataatcacacgctggttaataat--atgggaacgctatgt 1367

```

```

Db 18437 AATATATTTAAATTAATTAATTAATTTTAAATAATTTCTTAATGATATTATTTAT 18496
Oy 1368 aagatcacgcgaataatcctcttggtgcaacaataaacaagaagaaga 1427
Db 18497 AAAAAATATTTATATATATATATATATATATATATATATATATATATATAT 18556
Oy 1428 gaacacgatcttcttgatccatcgaatgatacgaatgatacgaatgatac 1487
Db 18557 ATAAATTTATATATATATATATATATATATATATATATATATATATATAT 18616
Oy 1488 agttcgaagtcctcgaagcgtgtaaccatcgaactgaactgaactgcttct- 1546
Db 18617 AAAAAATATTTTATATATATATATATATATATATATATATATATATATAT 18676
Oy 1547 tgcattcttaacatcttcttgatgtaagtaagtaagtaagtaagtaagtaag 1606
Db 18677 TATATATATATATATATATATATATATATATATATATATATATATATAT 18736
Oy 1607 aatattatattagatagatagatagatagatagatagatagatagatagat 1666
Db 18737 AATTTAAATATGATATATATATATATATATATATATATATATATATATAT 18796
Oy 1667 tcaactgtcttctcctcctcgaatgataacatcgaatgataacatcgaat 1726
Db 18797 TTTAATTTTAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 18856
Oy 1727 tcttctcgaagtcgaagtaagtaagtaagtaagtaagtaagtaagtaag 1786
Db 18857 ATATTTGATATATATATATATATATATATATATATATATATATATATAT 18915
Oy 1787 ttactacatagatagatagatagatagatagatagatagatagatagat 1846
Db 18916 TTTAATATATATATATATATATATATATATATATATATATATATATAT 18975
Oy 1847 gatagtcgaatcttctcgaatgataacatggaatggaatggaatggaatg 1906
Db 18976 AATTAATATATATATATATATATATATATATATATATATATATATATAT 19035
Oy 1907 aacagctggaatcttctcgaatgataacatggaatggaatggaatggaat 1966
Db 19036 AATATATATATATATATATATATATATATATATATATATATATATAT 19095
Oy 1967 aacatattatataacatgatacgaatgatacgaatgatacgaatgatac 2026
Db 19096 TTTTAAATATATATATATATATATATATATATATATATATATATATAT 19155
Oy 2027 ttatattagatagatagatagatagatagatagatagatagatagat 2086
Db 19156 TTAATATATATATATATATATATATATATATATATATATATATATAT 19215
Oy 2087 attaatcttctcgaatgataacatgatacgaatgatacgaatgatacga 2146
Db 19216 ATATATATATATATATATATATATATATATATATATATATATATATAT 19275
Oy 2147 ttcacacacacacacacacacacacacacacacacacacacacacacac 2206
Db 19276 TAAATATATATATATATATATATATATATATATATATATATATATAT 19335
Oy 2207 taacgatagacacacacacacacacacacacacacacacacacacacac 2266
Db 19336 TATATATATATATATATATATATATATATATATATATATATATATAT 19395
Oy 2267 taacacacacacacacacacacacacacacacacacacacacacacac 2326
Db 19396 AAAAAATTTCTTAAATGATATATATATATATATATATATATATATATAT 19455
Oy 2327 atgcacggtgcgacacacacacacacacacacacacacacacacacacac 2386
Db 19456 TTTTAAATATATATATATATATATATATATATATATATATATATATAT 19515

```

RESULT 12

```

PFMALIP3/C
LOCUS        PFMALIP3 67970 bp DNA INV 15-DEC-1999
DEFINITION   Plasmodium falciparum MALIP3, complete sequence.
ACCESSION    AL031746
VERSION      AL031746.9 GI:6594243
KEYWORDS     HTG.
SOURCE       malaria parasite P. falciparum.
ORGANISM     Plasmodium falciparum.
REFERENCE    1 (bases 1 to 67970)
AUTHORS      Bowman, S., Churcher, C., Harris, B., Harris, D., Lawson, D., Quail, M.
              and Barrell, B.
TITLE        Direct Submission
JOURNAL      Submitted (24-SEP-1998) P. falciparum Genome Sequencing Consortium,
              The Sanger Centre, Wellcome Trust Genome Campus, Hinxton, Cambridge
              CB10 1SA UK
COMMENT      On Dec 16, 1999 this sequence version replaced gi:5763807.
              For more information about this sequence or the Malaria Project,
              see http://www.sanger.ac.uk/Projects/P_falciparum. IMPORTANT: This
              sequence is unfinished and does not necessarily represent the
              correct sequence. Work on the sequence is in progress and the
              release of this data is based on the understanding that the
              sequence may change as work continues. The sequence may be
              contaminated with foreign sequence from E.coli, yeast, vector,
              phage etc.

FEATURES
    source
        1..67970
            /location="Qualifiers"
            /organism="Plasmodium falciparum"
            /strain="3D7"
            /db_xref="taxon:5833"
            /chromosome="1"
            /complement(join(1748..2598,2748..2848,2990..3276))
            /gene="MALIP3.01"
            /complement(join(1748..2598,2748..2848,2990..3276))
            /gene="MALIP3.01"
            /note="MALIP3.01, conserved hypothetical protein, len: 412
              aa, similarity: UF0006 family eg to
              YBL055C/YBL0512/YBL0511, YB5_YEAST (418 aa), fasta
              scores: opt: 316, E(): 1.1e-12, (33.2% identity in 271 aa
              overlap)"
            /codon_start=1
            /product="conserved hypothetical protein, UPP0006 family"
            /protein_id="CAB63556.1"
            /db_xref="GI:6594244"
            /translation="MKLVFHYIKYINLVFYISIFLKSNSLKIYNDLRYSTVNVKKY
              LQIKRSNLKKNHNIKMEWNESSFIQISGLTQKMPGYNVSKKHEDNLQNVNRAK
              NNANDKIITCTCLAEIDSKICETIDPREKPIYLSGVPTNCYETIDKRNKHEKE
              ILAKREBEFKIRNEOVENSKEKNGKKICDGEKDMNNNEILLENLDTIFGFKY
              NEKDEYLENFKNIKIPNRIIVGIGELGIDRLYFCSKYIQIKYIFFOKLQVFN
              LPMLHMCNCESEFEKVIDIYKFLPEKNGGVYHSFTQEDYIVHIVQYKNLYIGVNG
              CSLSLEININAVKKIPLNLLETDAPCGVKTHASVEYIKDYERKRAYTNLKKIKN
              IIKDDNTIFKERNRPYNIA"
            /complement(2599..2610)
            /note="potential splice acceptor sequence"
            /complement(2742..2747)
            /note="potential splice donor sequence, atg/gttaa"
            /complement(2849..2861)
            /note="potential splice acceptor sequence"
            /complement(2984..2989)
            /note="potential splice donor sequence, aaa/gtaaaa"
            /gene="MALIP3.02"
            /gene="MALIP3.02"
            /note="MALIP3.02, hypothetical protein, len: 163 aa,
              contains possible signal sequence"
            /codon_start=1
            /product="hypothetical protein, MALIP3.02"
            /protein_id="CAB63557.1"
            /db_xref="GI:6594245"
            /translation="MKLLNRFVVLCPILIFFLNSVGLGNRRNNINHEHETNAK
              AMRLLSGEINSLKLDGDEIKILNDEKHDSYKWKSYSPISNLEEKQSQIDLR
              KQDINENATKIIEDROEFYILNDEIENIATREVLNNEDELYIOSFKSLIDIOS

```

[illegible]

44.18; Pred. No.

1101 cgaataatagcttgaagaaccccttagaanaattllggaattaccggttgcattgtaataata 1160
 || || | | ||| |||| | | || || | |
 2825 CTATTAAATTATAATTACGATATACGTTTTTTTTAAAAAAAATAATATTTTTTAAAAAA 2766

| | | | |
|----|------|---|------|
| Oy | 1161 | gattgctgcygaacaaatctgttaacaaatcagctgctgaacaaatcactaaccttcac | 1220 |
| Db | 2765 | ATTTTATTTTAAAAATGAAAAATTAATTAATTTATTTATTAATAATTTATTTATTA | 2706 |
| Oy | 1221 | agaaaaaacaacttaagagaagcttaacatcaccatatacctggtcactacccctac | 1280 |
| Db | 2705 | AAATTTTTTGTGTTATTTTAAAAAACATGATTTTATTTAATAAATATTTTATTA | 2646 |
| Oy | 1281 | gtatgcatactagagacctaagaatcagttatcagctgctgcgaataatgaatccacgc | 1340 |
| Db | 2645 | TAAATACATTTAGAAATTTTAAAAATTTATATTAATTTATTAATTAATTTATTTT | 2586 |
| Oy | 1341 | gtggttaaatctatgagagccgtatcgtacgacacgcgaatcactcctcgtgcga | 1400 |
| Db | 2585 | CTATATATATATATATATTTATTAATATTTCAATATATTAATTTATTAATTAATA | 2526 |
| Oy | 1401 | acaaataaacaacaacagaagaaaaaagaacagatcttcctggtacatccaagat | 1460 |
| Db | 2525 | TTAATTTAAATATATTAATAAAAAAATAAAAAAATAATTTTATTTATTAATTAAT | 2466 |
| Oy | 1461 | ctaaatgcatagactctccttggtctcagcttcagtgacgcctccacaagcgttaacc | 1520 |
| Db | 2465 | TAGTAATTAATTAATTTATTTATTAATTAATTAATTAATTTATTTATTAATTAAT | 2406 |
| Oy | 1521 | tgcacataaactcgtcctcttaatcgcacttacaatctatcgtctgaagat | 1580 |
| Db | 2405 | ATTAATATATGAAATGATTTTATTTATTTATTTATTTATTTATTTATTTATTTAT | 2346 |
| Oy | 1581 | taataa-gagcgaactcgttaacatcacaatttatc-tgaatacgtatgtagtatc | 1638 |
| Db | 2345 | TAAAAATTTATTTTCATTTATTAATATATATATATATATATATTAATTAATTTAT | 2286 |
| Oy | 1639 | tcacaata-----catccttgagctcttaacactaactcgtcttcctccacgta | 1690 |
| Db | 2285 | TTTAATTAATTTATTTATTAATTAATTAATTTATTAATAAATAGTTATTAAGATTA | 2226 |
| Oy | 1691 | taaatataatcatcagagcgttaaaaaagcttctcctatcttcgcgagatgaagata | 1750 |
| Db | 2225 | ATTAATTAATTTTAAAAAATAATTTTAAAGTTTAAATTAATTAATAATTAATTA | 2166 |
| Oy | 1751 | aacctaatgacttaactcttgaanaatgaaccccttaacactagatctaactcgt | 1810 |
| Db | 2165 | TGAAATAGGGGAATAATTTATTTTCATTTTTCATATATATATATATATATATAT | 2106 |
| Oy | 1811 | tgcttctgctgcatacgaacagccctcaacactgtagatcgaactcttcctgaata | 1870 |
| Db | 2105 | ATTAATTAATTTTCAGATTTTACTGATTTAAATTAATTAATTTATTAATTAATTA | 2046 |
| Oy | 1871 | ttaaatcaggaatcaactgctactcaactagaagaacagctgagatctaacttat | 1930 |
| Db | 2045 | TTGAATTTAATTAATTAATATGATATATATATATTAATAATATATGAATTTGA-- | 1988 |
| Oy | 1931 | ctaaagacaatcttgaanaatgltataactctcaacatatacttaanaatatyagc | 1990 |
| Db | 1987 | ATCATTTTAAATTTTATTTATTTATTAATAAATATTTTATTAATTTAAATTTTAT | 1928 |
| Oy | 1991 | ctaaatgactccactgctccttaaatcttcttatactagttataaact | 2030 |
| Db | 1927 | TTATTAATAATTAATTTATTAATAAATAGTTTATTAAGTAATTAATTAATAATTA | 1868 |
| Oy | 2051 | atgaacaaataatagctggtgaactcaaacatccacttaaatcttgcgaactcaaa | 2110 |
| Db | 1867 | TTTAATAAATAAATTTTATTTTAAGTTTATTTATTAATTAATAA--TTTATGAAT | 1811 |
| Oy | 2111 | atctataatctagcacaataacaatgcataagaagctccaaaaaaactctgttaacg | 2170 |
| Db | 1810 | ATTAATTTATTTTCATTTTATTTTATTAATATTAATATATATTAATAATTTATTC | 1751 |
| Oy | 2171 | aaactccaactcttcttcttctgaagcaagaataacagatagaacactatctt | 2230 |
| Db | 1750 | TAGCATTAATAATAATTAATTTATTTATTAATTAATTAATTTATTAATTTGAATA | 1632 |

| | | | |
|------------|---|--|------------------------------|
| Oy | 2231 | gtcgaaatggaagtagtaatacaacaaatgcgaattctttaaataatcagaaccatcgc | 2290 |
| Dd | 1693 | -----TATATATATATATATATATATATAAATAATGAATGTAATTTTAAAAAATTATTTTTAA | 1638 |
| Oy | 2291 | cgcctcaagaatctgtatcatcagtagaggltgatcaataacatgcacgtgcgattccagaattcg | 2350 |
| Dd | 1637 | TTTTTATATATATAAATAAATTTCTTTATTAATTAATTTTAAAAAATTATTTATTAATAATA | 1578 |
| Oy | 2351 | gacaacatggaaaecggagatcaaatatctaacttaataataataataaatttgagtaat | 2410 |
| Dd | 1577 | ATTATTTATATAAAAATAAGTTTATTAAAGTAAATTTAATTAATTAATTAATTTTAAAAAAA | 1518 |
| Oy | 2411 | gtgtttcttcgacctatgcgggccaacaaaa | 2441 |
| Dd | 1517 | ATATTTTAAAGTTTAAATTAATTAATAATAA | 1487 |
| RESULT_14 | | | |
| DMU37541/c | | | |
| LOCUS | DMU37541 | 19517 bp | DNA circular INV 02-MAR-2001 |
| DEFINITION | Drosophila melanogaster complete mitochondrial genome. | | |
| ACCESSION | U37541 | | |
| VERSION | U37541.1 | GI:1166529 | |
| KEYWORDS | fruit fly, | | |
| SOURCE | Mitochondrion Drosophila melanogaster | | |
| ORGANISM | Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta; | | |
| | Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; | | |
| | Muscomorpha; Ephydroidea; Drosophilidae; Drosophila. | | |
| REFERENCE | 1 (bases 12511 to 12682) | | |
| AUTHORS | Clarry,D.O., Goddard,J.M., Martin,S.C., Fauron,C.M. and Wolstenholme,D.R. | | |
| TITLE | Drosophila mitochondrial DNA: a novel gene order | | |
| JOURNAL | Nucleic Acids Res. | 10 (21), 6619-6637 | (1982) |
| MEDLINE | 83090428 | | |
| REFERENCE | 2 (bases 5269 to 5695) | | |
| AUTHORS | Clarry,D.O., Wahlteitner,J.A. and Wolstenholme,D.R. | | |
| TITLE | Transfer RNA genes in Drosophila mitochondrial DNA: related 5' flanking sequences and comparisons to mammalian mitochondrial tRNA genes | | |
| JOURNAL | Nucleic Acids Res. | 11 (8), 2411-2425 | (1983) |
| MEDLINE | 83220794 | | |
| REFERENCE | 3 (bases 404 to 5272) | | |
| AUTHORS | de Bruijn,M.H. | | |
| TITLE | Drosophila melanogaster mitochondrial DNA, a novel organization and genetic code | | |
| JOURNAL | Nature | 304 (5923), 234-241 | (1983) |
| MEDLINE | 83245048 | | |
| REFERENCE | 4 (bases 804 to 1778) | | |
| AUTHORS | Satta,Y., Ishiya,H. and Chiguza,S.I. | | |
| TITLE | Analysis of nucleotide substitutions of mitochondrial DNAs in Drosophila melanogaster and its sibling species | | |
| JOURNAL | Mol. Biol. Evol. | 4 (6), 638-650 | (1987) |
| MEDLINE | 88174373 | | |
| REFERENCE | 5 (bases 5268 to 13619) | | |
| AUTHORS | Garesse,R. | | |
| TITLE | Drosophila melanogaster mitochondrial DNA: gene organization and evolutionary considerations | | |
| JOURNAL | Genetics | 118 (4), 649-663 | (1988) |
| MEDLINE | 88212147 | | |
| REFERENCE | 6 (bases 441 to 2967) | | |
| AUTHORS | Satta,Y. and Takahata,N. | | |
| TITLE | Evolution of Drosophila mitochondrial DNA and the history of the melanogaster subgroup | | |
| JOURNAL | Proc. Natl. Acad. Sci. U.S.A. | 87 (24), 9558-9562 | (1990) |
| MEDLINE | 91088557 | | |
| REFERENCE | 7 (bases 14215 to 14512) | | |
| AUTHORS | Ballard,J.W., Olsen,G.J., Faith,D.P., Odgers,W.A., Rowell,D.M. and Atkinson,P.W. | | |
| TITLE | Evidence from 12S ribosomal RNA sequences that onychophorans are modified arthropods | | |
| JOURNAL | Science | 258 (5086), 1345-1348 | (1992) |
| MEDLINE | 93088057 | | |

```

REFERENCE      8 (bases 14917 to 19517)
AUTHORS        Lewis,D.L., Farr,C.L., Farquhar,A.L. and Kaguni,L.S.
TITLE          Sequence, organization, and evolution of the A+T region of
               Drosophila melanogaster mitochondrial DNA
JOURNAL        Mol. Biol. Evol. 11 (3), 523-538 (1994)
MEDLINE        94285822
REFERENCE      9 (bases 1 to 408; 13319 to 19517)
AUTHORS        Lewis,D.L., Farr,C.L. and Kaguni,L.S.
TITLE          Drosophila melanogaster mitochondrial DNA: completion of the
               nucleotide sequence and evolutionary comparisons
JOURNAL        Insect Mol. Biol. 4 (4), 263-278 (1995)
MEDLINE        96423163
REFERENCE      10 (bases 1 to 19517)
AUTHORS        Lewis,D.L., Farr,C.L. and Kaguni,L.S.
TITLE          Direct Submission
JOURNAL        Submitted (03-OCT-1995) Laurie S. Kaguni, Biochemistry Department,
               Michigan State University, East Lansing, MI 48824-1319, USA
FEATURES       source
               1. 19517
               /organism="Drosophila melanogaster"
               /organelle="mitochondrion"
               /db_xref="taxon:7227"
               /note="derived from new and previously submitted
               sequences; sequence is a composite containing sequences
               obtained from different Drosophila melanogaster strains"
               1. 65
               /product="tRNA-Ile"
               complement(97..165)
               /product="tRNA-Gln"
               171..239
               /product="tRNA-Phe"
               240..1265
               /codon_start=1
               /transl_table=5
               /product="NADH dehydrogenase subunit 2"
               /protein_id="AAC47811.1"
               /db_xref="GI:1166530"
               /translation="MPNNSKSLFTITMIGLITVTSNMGAMGELNITLSPIL
               LSDNNNMSTESLAKFELQVASTVLESSLIMAKNNNEINSPMTIMSL
               LKSGAPFHPFPMNMEGLTMMALMTWKIAPMLITSYNIKTLISVLTSLI
               KAIIGLNOTSLRKLKLAFFSINHLGWLSSIMTSESLILFFPFSLSLVLFMEFI
               KLFHLNOLFMEVNSKILKFTLFMNFSLISGGLPPGLPKMLVIOQLTLCQYEMLT
               IMMSTLITLFFLYLRICYSAFMNNYFENNMIKMNNSINYNMYMIMTFEFSIFGLFLI
               SLFVFMF"
               1264..1329
               /product="tRNA-Trp"
               complement(1322..1383)
               /product="tRNA-Cys"
               complement(1403..1468)
               /product="tRNA-Tyr"
               join(1470..1472,1474..3009)
               /note="mechanism underlying reading frame shift after
               first codon uncertain"
               /codon_start=1
               /transl_table=5
               /product="cytochrome c oxidase subunit I"
               /protein_id="AAC47812.2"
               /db_xref="GI:7412849"
               /translation="MSRWLFSTNHRDICTLYFICGAMGVSTSLIRAELEGHPG
               ALIGDOIYNIIVTAHAFFIMFPMWPIIMIGFNMVLVPLMGAPMAPPRMNMSEFM
               LIPALSLILSVSNGAGTGMTVYPPISAGIAGGASVDLAIFSLIAGISLIGLA
               VNFITVIMRSTGTSIDRMLPVMVSIVTALLILSLPLAGATIMLTDRINLSE
               FPPAGGGPILIOHLFHPFGHEVYIILIPGFMISHIISQESGKETGSGMTIYAM
               LAIGLGIIVMAHMEYVGMVDYTRAFYTSATMIJAVPGIKIFSWLATLHGTOISY
               PAILMAIGVLFLETVGLTGVVLANSSVDIILHDYVYVAHFHYVLSMGAVPAIMAGF
               IHMYPLETGLTNNKMLKSHFTIMFGVNLIFPPQHFLLAGMPRYSYDPAYTTWM
               IYSTGISTSLIGLILFFFTIMESLVSQROYIVPIQLNSIEMVQMPAHSYSPLP
               ILFN"
               3012..3077
               /product="tRNA-Leu"
               3083..3767
               /note="tRNA stop codon is completed by the addition of 3' A
               residues to the mRNA"

```

```

               /codon_start=1
               /transl_except=(pos:3767,aa:TERM)
               /transl_table=5
               /product="cytochrome c oxidase subunit II"
               /protein_id="AAC47813.1"
               /db_xref="GI:1166532"
               /translation="MSTWANTGLQDSASPLMEQLIFPHDHALITLWITVLGYLWM
               LFENNVNREFLHGLIEMITLIPAILFTALPRLRLYLDEINPESVTLKSI
               GOMVSEYSDENNIEPDSYMIPTNEMDGGPELLDVDRNVYPMNSQIRIIVTADVI
               HSWTPALGVKVDGTGRLNQTNFTNREGLFYGGCSEICGANHSFMPVIESVYV
               FTKWISSNNS"
               3768..3838
               /product="tRNA-Lys"
               3840..3906
               /product="tRNA-Asp"
               3907..4068
               /codon_start=1
               /transl_table=5
               /product="Argase 8"
               /protein_id="AAC47814.1"
               /db_xref="GI:1166533"
               /translation="MPQMAPISMLLFIITPILFCSINYSTPMSPKSNELNI
               NLSNMWKW"
               4062..4736
               /codon_start=1
               /transl_table=5
               /product="Argase 6"
               /protein_id="AAC47815.1"
               /db_xref="GI:1166534"
               /translation="MWTNLFVFDPLAIFNPSLNLSTFLGLMTPSIYWLMPRYNI
               MNNSILTLHKEEFTLGPSSNGNSSTFIFISLSLIFNNFGLPFIYFTSHTLT
               LSLALPILCMFLMGWINHTOHMFAHLVPOGPAIIMPVCIETISNIIIRGTLAVR
               LTNMAGHLITFLIGNTSSSYMLMTLMAQIALVLVESAVMIGSYVAVASTL
               YSEEVN"
               4736..5524
               /codon_start=1
               /transl_table=5
               /product="cytochrome c oxidase subunit III"
               /protein_id="AAC47816.1"
               /db_xref="GI:1166535"
               /translation="MSTHSNHPFHLVDYSPWPLTGAIGAMTVYSGWVKWPHOYDISL
               FVGNITITLVYQWMDVDSREGTYOGLHTYAVTIGLRGMILFIISEVLFVSEFAF
               FHSLSPLAIEIGASMPMGITSPNPQIOLMTATILLASGVYTAHHSILMENNHSOT
               TQGLFPTVLTGTYFTTIOAYEYIEAPETIADSYGTFPMATGFGHIVHIGTTFILV
               CLURLHNHPSKNHFGFEPAAMVHFVYVWVLFYITTYWNG"
               5543..5607
               /product="tRNA-Gly"
               5608..5961
               /codon_start=1
               /transl_table=5
               /product="NADH dehydrogenase subunit 3"
               /protein_id="AAC47817.1"
               /db_xref="GI:1166536"
               /translation="MFSIIFALLILLITTYIVFLASILSKKALIDREKSPPEC
               GFDPKSSRLPSLRFPLITITLIFDFVEIALILPWIIMKYSNMIMWTITSIFILLILI
               GLYHFNQGMKNSN"
               5981..6045
               /product="tRNA-Ala"
               6055..6118
               /product="tRNA-Arg"
               6119..6183
               /product="tRNA-Asn"
               6184..6251
               /product="tRNA-Ser"
               6252..6318
               /product="tRNA-Glu"
               complement(6337..6401)
               /product="tRNA-Phe"
               complement(6401..8124)
               /note="tRNA stop codon is completed by the addition of 3' A
               residue to the mRNA"
               /codon_start=1
               /transl_except=(pos:complement(6401..6402),aa:TERM)

```


CDS complement(join(31966, .32476,32675, .32775))
 /gene="MALP3.04"
 /note="MALP3.04, conserved hypothetical membrane protein, len: 203 aa, similarity: P. falciparum chromosome 2, PB0110W, 096126 predicted integral membrane protein (255 aa), fasta scores: opt: 335, E(): 4.9e-15, (36.1% identity in 191 aa overlap)"
 /codon_start=1
 /product="conserved hypothetical membrane protein, MALP3.04"
 /protein_id="CAB63559.1"
 /db_xref="GI:6594247"
 misc_feature translation="MKKSYTFINTVLTLLFLTLFLTYNYNDPFSKRFKNKKIKDIN
 RKRIIAASEOKYTWEDPCLINBELRLPHNDSPYLPKHYETIDINELISTNS
 TKIMKETKKMOKNETKDMNNHNRWDEPFMHYKMAIYLYKRLINILKDTLPTDI
 HDKETITWIKIODEYLFPIYLOVEMVRLITLLEFPYNNKE"
 complement(32477, .33486)
 misc_feature /note="potential splice acceptor sequence"
 complement(32669, .32674)
 misc_feature /note="potential splice donor sequence, aaa/gtatac"
 join(36657, .36743,36864, .37343)
 /gene="MALP3.05"
 CDS join(36657, .36743,36864, .37343)
 /gene="MALP3.05"
 /note="MALP3.05, hypothetical protein, len: 188 aa"
 /codon_start=1
 /product="hypothetical protein, MALP3.05"
 /protein_id="CAB63560.1"
 /db_xref="GI:6594248"
 misc_feature translation="MRIKMSGIFPIKLLISPCYFPCFNCKMISYRKDLIWSN
 CENYSIDSLAAGSESEKERYKQIDNELKLSINITYEEKKKEITVSGFDLNNIN
 IANELKINHSTNIENIFSLINDSSKSPVLTPLKEGSIIPMLNNVRLPLDLCQ
 RNPRLIIIGLIVLIALYVFENFNFC"
 36744, .36749
 misc_feature /note="potential splice donor sequence, aag/gtatga"
 36854, .36863
 misc_feature /note="potential splice acceptor sequence"
 complement(join(38049, .39995,40210, .40284))
 /gene="garp"
 CDS complement(join(38049, .39995,40210, .40284))
 /gene="garp"
 /note="MALP3.06, garp, len: 673 aa, similarity: almost
 identical to GARP_PLAF6 (678 aa), fasta scores: 97.6%
 identity in 678 aa overlap"
 /codon_start=1
 /product="hypothetical garp protein"
 /protein_id="CAB63561.1"
 /db_xref="GI:6594249"
 misc_feature translation="MNVLFISYNICILFPYVCITLSTKCFNSGLKNILNKSPDS
 ITGLRLNLTLEKNDKSKSETLKEKKEKEDVPPTTSDNLIKNAHNNNEISSDP
 TIIIVNWDKNSYDKKDKKKEKKKKKKKKKKKKKKKKKKKKKKKKKKKKKK
 DKKKEENSEVSLYTKTGQHKRNATKEGLENLYEEMVLTINNAGGILLSSPYORE
 OGCGCISVYETSNDTKDNKENSIDKEDKQKEMKKKITEKKKKKKKKKKKK
 KTEKKKKOEKK
 HKKEENHEEPDQDTSEFTNNEIWPPLSPDTDTTDEEKKKEEKEEBENHKK
 EHKKEENHEEKKKEENKSHKSKGDKGKKKKKKKKKKKKKKKKKKKKKKKK
 DKDGEVILNDEKACEONHTVESPLOSCGLIDEPDLTLMDSKYEKNLSIO
 BOLIGTGRVAVPRNRHKKKKKTEALQKOKHYDKEDKKKEKKEEPEKSYVO
 EDDEVEDE
 DDDEDEDDDDDEDEDEDEDEDEDEDEDEDEDESEKKIRNLNKKAKTI"
 complement(39996, .40005)
 misc_feature /note="potential splice acceptor sequence"
 complement(40204, .40209)
 misc_feature /note="potential splice donor sequence, aag/gtatac"
 join(45401, .46396,46562, .50233)
 /gene="MALP3.07"
 CDS join(45401, .46396,46562, .50233)
 /gene="MALP3.07"
 /note="MALP3.07, hypothetical protein, len: 1555 aa"
 /codon_start=1
 /product="hypothetical protein, MALP3.07"
 /protein_id="CAB63562.1"
 /db_xref="GI:6594250"

```

Query Match          3.7%: Score 118; DB 96; Length 67970;
Best Local Similarity 43.8%: Pred. No. 7.3e-07;
Matches 1011; Conservative 0; Mismatches 1265; Indels 34; Gaps 10;

/translation="MEOIGJANNIIFTEIGKNIIVRSQFPEIWRKKPLFEESKPLTISNSI
NIENCKEKKRCKRKNINMGDDKEEFVSRRKFVYERISLNNKPEKAIINTLKNTT
KTCLSRKKRIFKEIRINMGLEDYIDAIINDIKOPKNLYCSENYIDIKKWLNMKEY

QY 108 ataataaataataaatcattcaataaataattacaagaaataatcttaccatigt 167
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 7975 AATATATATTAATATATATTTCTTATTAATAAATTAATTAATAGACACATTTATTAATAA 8034
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 168 attgtataacaaatatctctctgttgatctggaataatactggaagttggaattata 227
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 8035 ATATATATTAAGAAATATTTTATCTTTTGATATATATATATATATTTAANTGCAATTAAT 8094
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 228 ataataaagaaataacgcattccatttggcttgatcacaggttaagtttcttgcttc 287
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 8095 TAAATTAATTAATAATATATAAATAATATATATTAATATATTAATTAACAAATATATATTA 8154
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 288 ctctgttatactatctgaaatccaaagagatctgattgaaagctgbaacatactc 347
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 8155 TATATTAATTAATAAATAATATATATTAATAATAA---TAAATTAATAAATAATTAATAA 8210
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 348 gttatgaccccccaaaaaaaaaaaaaaacaacaacaaccccccccgatagt 407
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 8211 AATATTAATTAATTAATATATATTAATTAATAAATTTAATAATATATTAATTAATAATAT 8270
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 408 ctggtcttgattgaagtttattcttgatcacaattacatgcattctcttgatctag 467
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 8271 AATTAATTAATAAATAATATATATTAATTAATAAATTAATAAATTAATAATATATTAATTA 8330
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 468 aagattctcttacaataaataatcgaattcatactcttgattataaataaatacg 527
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 8331 ATTAATATATATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 8390
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 528 agtfgaagat-alcogttatacgcacccccaacaaatgatttgattcttgctgaacc 585
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 8391 TATATTAATTAATTAATATATTAATTAATAAATAATTTATTAATTAATTAATTAATAATAT 8450
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 586 agcaaatcttacaagagatctggaataaacaacgaataaagaagaagaaagtagt 645
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 8451 ATTTAAATATTAATTAATTAATTAAT--ATATATATTAATTAATTAATTAATTAATTAAT 8508
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 646 gaccatgagagatctgaaataatlcacaagaagaaataagagatgacaaacaaagcttct 705
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 8509 AATATTAATTAATTAATTAATTAATAAATAATATATTAATTAATTAATTAATAATTAATTA 8568
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 706 gaaataaagtgccctgcacagcttctctccacaacaaataatcgacctatttgattctct 765
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 8569 AATATTAATTAATTAATTAATTAATTAATTAATTAATAAATAATTTATTAATTAATAATTA 8628
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 766 ggaattctgtaaaatttgcgaataacgatttggaataaatttattctgttagcgcacct 825
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 8629 TAATTAATTAATTAATAAATAAATAATTAATTAATTAATAAATAAATAAATAAATAAATAAT 8688
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 826 caatatatgctccaggttaattgcataaactctgtttaaagacattctgcttctct 885
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 8689 AATATTAATTAATTAATTAATAAATAAATAATTAATTAATTAATTAATAAATAAATAATTA 8748
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 886 ttgttctgcttctcttaactataatactcgcgataataatgatacaaalgtatalatacaca 945
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 8749 TTAATAATATATTAATTAATTAATTAATAAATAATTAATTAATAATTAATAATTAATTAAT 8802
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 946 aaaaatagctcggggaacatttctgaaataacttcttccaaacttcaagggaacactgac 1005
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 8803 GAAATATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 8862
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 1006 tcgaccttaaaatacgaatttcaagcgctgcagcttgagattactagcacaagaacataa 1065
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 8863 ATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 8922
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 1066 aggaccgcttcaagctatttacaagaatttccaaactgaatatagcttgaaactccttag 1125

```

[illegible]

| | | | |
|----|-------|--|-------|
| Db | 10000 | AATATTATATATTATTAAATAAATTAATAATATTTTACTATATTAAATTAATTAATTTTA | 10058 |
| Qy | 2189 | ttttatcggaacagaataaacaagatgagaaactttgtctgbygaatgggaagcagtaa | 2248 |
| Db | 10060 | TATTTTAAATTAATAATATATTTTATTAATAATAAATTAATTTTTCATATATTAATTT | 10119 |
| Qy | 2249 | tatacatlaagcaaatlttlaaaaaatlatataagccatacgcgctcaaaagatgttacc | 2308 |
| Db | 10120 | AATATTAATTAATAATATTTATTAATGCAAAATTAATTTTAAATTAATTAATTAATTTATG | 10179 |
| Qy | 2309 | tagttagtgcataataataatgcacgtgcgcatccgaattggagacaacaatgaacaacgga | 2368 |
| Db | 10180 | TATTTATATTTATTAATAAATTAATTTTTTGTGTGTAATAATTAATTAAGATTTTATATA | 10239 |
| Qy | 2369 | atlaaatatlaactttaaataaataaa | 2398 |
| Db | 10240 | TTGATGTATTATTTTATAAATTAATAAATAA | 10269 |

Search completed: October 17, 2001, 19:26:38
Job time: 16993 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd

OM nucleic - nucleic search, using sw model

Run on: October 17, 2001, 14:53:20 ; Search time 431.79 Seconds

4656.294 Million cell updates/sec

Title: US-09-502-426-1_COPY_1_3202

Sequence: 1 atgtggtattatattgtt...gagagagagaactagctcc 3202

Scoring table: IDENTITY_NUC

Searched: 730101 seqs, 313950809 residues

Total number of hits satisfying chosen parameters: 1460202

Maximum DB seq length: 200000000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database : N_Geneseq_0601: *

| | |
|-----|--|
| 1: | /SDSI1/gcgdata/gemseq/gemseqn/NA1980.DAT * |
| 2: | /SDSI1/gcgdata/gemseq/gemseqn/NA1981.DAT * |
| 3: | /SDSI1/gcgdata/gemseq/gemseqn/NA1982.DAT * |
| 4: | /SDSI1/gcgdata/gemseq/gemseqn/NA1983.DAT * |
| 5: | /SDSI1/gcgdata/gemseq/gemseqn/NA1984.DAT * |
| 6: | /SDSI1/gcgdata/gemseq/gemseqn/NA1985.DAT * |
| 7: | /SDSI1/gcgdata/gemseq/gemseqn/NA1986.DAT * |
| 8: | /SDSI1/gcgdata/gemseq/gemseqn/NA1987.DAT * |
| 9: | /SDSI1/gcgdata/gemseq/gemseqn/NA1988.DAT * |
| 10: | /SDSI1/gcgdata/gemseq/gemseqn/NA1989.DAT * |
| 11: | /SDSI1/gcgdata/gemseq/gemseqn/NA1990.DAT * |
| 12: | /SDSI1/gcgdata/gemseq/gemseqn/NA1991.DAT * |
| 13: | /SDSI1/gcgdata/gemseq/gemseqn/NA1992.DAT * |
| 14: | /SDSI1/gcgdata/gemseq/gemseqn/NA1993.DAT * |
| 15: | /SDSI1/gcgdata/gemseq/gemseqn/NA1994.DAT * |
| 16: | /SDSI1/gcgdata/gemseq/gemseqn/NA1995.DAT * |
| 17: | /SDSI1/gcgdata/gemseq/gemseqn/NA1996.DAT * |
| 18: | /SDSI1/gcgdata/gemseq/gemseqn/NA1997.DAT * |
| 19: | /SDSI1/gcgdata/gemseq/gemseqn/NA1998.DAT * |
| 20: | /SDSI1/gcgdata/gemseq/gemseqn/NA1999.DAT * |
| 21: | /SDSI1/gcgdata/gemseq/gemseqn/NA2000.DAT * |
| 22: | /SDSI1/gcgdata/gemseq/gemseqn/NA2001.DAT * |

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB | ID | Description |
|------------|-------|-------------|--------|----|-----------|---|
| 1 | 3202 | 100.0 | 6888 | 21 | AAFA59599 | DNA encoding a cytochrome b5 domain protein |
| c 2 | 251 | 7.8 | 936 | 22 | AAFA58252 | Oligonucleotide D1 |
| c 3 | 251 | 7.8 | 936 | 22 | AAFA58254 | Oligonucleotide D1 |
| c 4 | 251 | 7.8 | 936 | 22 | AAFA58257 | Oligonucleotide D1 |
| c 5 | 251 | 7.8 | 936 | 22 | AAFA58259 | Oligonucleotide D1 |
| c 6 | 251 | 7.8 | 936 | 22 | AAFA58262 | Oligonucleotide D2 |
| c 7 | 251 | 7.8 | 936 | 22 | AAFA58255 | Oligonucleotide D2 |
| 8 | 250.6 | 7.8 | 936 | 22 | AAFA58252 | Oligonucleotide D1 |
| 9 | 250.6 | 7.8 | 936 | 22 | AAFA58254 | Oligonucleotide D1 |
| 10 | 250.6 | 7.8 | 936 | 22 | AAFA58257 | Oligonucleotide D1 |
| 11 | 250.6 | 7.8 | 936 | 22 | AAFA58259 | Oligonucleotide D2 |

| | | | | | | | | |
|---|----|------|-----|-------|----|-----------|--|---------------------|
| C | 19 | 74 | 2.3 | 2104 | 13 | AAQ25273 | | Sequence encoding |
| C | 20 | 73 | 2.3 | 1864 | 8 | AAW71405 | | Sequence of ANS-1 |
| C | 21 | 70.8 | 2.2 | 8310 | 20 | AAZ29911 | | cDNA encoding a SC |
| C | 22 | 69.6 | 2.2 | 4590 | 7 | AAN60472 | | Sequence encoding |
| C | 23 | 68.8 | 2.1 | 875 | 21 | AAA01920 | | Human colon cancer |
| C | 24 | 68.6 | 2.1 | 1864 | 8 | AAW71405 | | Sequence of ANS-1 |
| C | 25 | 68.2 | 2.1 | 2104 | 13 | AAQ25273 | | Sequence encoding |
| C | 26 | 67.4 | 2.1 | 5760 | 6 | AAN50530 | | Sequence encoding |
| C | 27 | 67.4 | 2.1 | 19124 | 18 | AAAT72882 | | Plasmodium var-7 g |
| C | 28 | 67.4 | 2.1 | 19124 | 21 | AAZ98287 | | Plasmodium var-7 p |
| C | 29 | 67.2 | 2.1 | 875 | 21 | AAA01920 | | Human colon cancer |
| C | 30 | 66 | 2.1 | 2486 | 21 | AAA97037 | | Nucleotide sequenc |
| C | 31 | 65.8 | 2.1 | 2486 | 21 | AAA97037 | | Nucleotide sequenc |
| C | 32 | 65.4 | 2.0 | 2503 | 15 | AAQ53480 | | PMPX30 xylanase CD |
| C | 33 | 65.4 | 2.0 | 20674 | 21 | AAQ58017 | | Arachidonic acid m |
| C | 34 | 65.2 | 2.0 | 6033 | 21 | AAA70152 | | Plasmodium falci |
| C | 35 | 65 | 2.0 | 2418 | 13 | AAQ27886 | | P.falciparum GBP13 |
| C | 36 | 64.4 | 2.0 | 5409 | 21 | AAA70151 | | Plasmodium falci |
| C | 37 | 64.2 | 2.0 | 1431 | 21 | AAZ37082 | | DNA sequence enco |
| C | 38 | 64.2 | 2.0 | 1671 | 13 | AAQ24134 | | 50 kb subunit of S |
| C | 39 | 63.6 | 2.0 | 8920 | 15 | AAQ62924 | | Carbamoyl-phosphat |
| C | 40 | 63 | 2.0 | 20674 | 21 | AAQ58017 | | Arachidonic acid m |
| C | 41 | 62.6 | 2.0 | 4818 | 18 | AAAT62177 | | Candida boidinii A |
| C | 42 | 62.6 | 2.0 | 4818 | 18 | AAAT79909 | | Actin gene. Candid |
| C | 43 | 62.6 | 2.0 | 9789 | 17 | AAAT41852 | | cDNA encoding Plas |
| C | 44 | 61.8 | 1.9 | 1864 | 15 | AAQ78892 | | Aspergillus nidula |
| C | 45 | 61.8 | 1.9 | 5852 | 12 | AAQ11710 | | Dictyostellium plas |

ALIGNMENTS

| | | |
|-----------|---|---------------------------|
| RESULT | 1 | |
| AAAA59599 | | |
| ID | AAAA59599 | standard; DNA; 6888 bp. |
| XX | | |
| AC | AAAA59599; | |
| XX | | |
| DT | 14-NOV-2000 | (first entry) |
| XX | | |
| DE | DNA encoding a cytochrome P450 enzyme designated DMF4. | |
| XX | | |
| KM | DMF4; cytochrome P450 enzyme; brassinosteroid; 22alpha-hydroxylation; | |
| RW | plant phenotype; cell elongation; ss. | |
| XX | | |
| OS | Arabidopsis sp. | |
| XX | | |
| FH | Key | Location/Qualifiers |
| FT | promoter | I..3203 |
| FT | | /+tag= a |
| FT | TATA_signal | 3060..3125 |
| FT | | /+tag= b |
| FT | CDS | 3203..6110 |
| FT | | /+tag= c |
| FT | | /product= "DMF4" |
| FT | | /note= "contains introns" |
| FT | exon | 3203..3423 |
| FT | | /+tag= d |
| FT | intron | 3424..3503 |
| FT | | /+tag= e |
| FT | exon | 3504..3828 |
| FT | | /+tag= f |
| FT | intron | 3829..3913 |
| FT | | /+tag= g |
| FT | exon | 3914..4056 |
| FT | | /+tag= h |

```

FT      Intron      4067..4164
FT      /tag= i
FT      Exon      4165..4479
FT      /tag= j
FT      Intron      4480..4631
FT      /tag= k
FT      Exon      4632..4724
FT      /tag= l
FT      Intron      4725..4815
FT      /tag= m
FT      Exon      4816..4894
FT      /tag= n
FT      Intron      4895..5000
FT      /tag= o
FT      Exon      5001..5110
FT      /tag= p
FT      Intron      5111..5864
FT      /tag= q
FT      Exon      5865..6110
FT      /tag= r
FT      3'UTR      6011..6468
FT      /tag= s
XX      WO200047715-A2.
XX      17-AUG-2000.
XX      11-FEB-2000; 2000WO-US03820.
XX      11-FEB-1999; 99US-0119657.
XX      11-FEB-1999; 99US-0119658.
XX      (ARIZ-) ARIZONA BOARD OF REGENTS.
XX      Azpiroz R, Choe S, Feldmann KA;
XX      WPI; 2000-549142/50.
XX      P-PSDB; AAB07921.
XX      New isolated dwf4 polynucleotide useful for altering the phenotype of
XX      plants, for diagnostic assays and in the production of antibodies -
XX      Claim 3; Fig 10A-G; 113pp; English.
XX      The present sequence encodes a DWF4 polypeptide. The polypeptide is a
XX      cytochrome P450 enzyme that mediates multiple steps in synthesis of
XX      brassinosteroids. Specifically, it mediates multiple
XX      22alpha-hydroxylation steps in brassinosteroid biosynthesis. The DWF4
XX      polynucleotide is used for altering the phenotype of a plant. DWF4
XX      plants display a dramatic reduction in the length of different organs,
XX      and this size reduction is attributable to a defect in cell elongation.
XX      The DWF4 polynucleotides and polypeptides can be used in diagnostic
XX      assays and to generate antibodies, which can be used to produce
XX      immunogenic compositions.
XX      Sequence 6888 BP; 2294 A; 1010 C; 1193 G; 2391 T; 0 other:
SQ

```

Query Match 100.0%; Score 3202; DB 21; Length 6888;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 3202; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

QY      181 aatctcatcttggatatagaagaataatgagcttggaaatttaataataaagaaga 240
DB      181 aatatcatcttggatatagagaaaaatgagcttggaaatttaataataaagaaga 240
QY      241 taatcgattccatttgggttgaatcacagtttaagtttggttcttcttggatata 300
DB      241 taatcgattccatttgggttgaatcacagtttaagtttggttcttcttggatata 300
QY      301 atatagttaatacaaaaagagatgataatgaaagtgaaacataatcttcgtatgacccca 360
DB      301 atatagttaatacaaaaagagatgataatgaaagtgaaacataatcttcgtatgacccca 360
QY      361 aaaaaaaaaaaaaacaacaacaaaccccccccgatatagtttggttcgatt 420
DB      361 aaaaaaaaaaaaaacaacaacaaaccccccccgatatagtttggttcgatt 420
QY      421 aggttatttgatcacaatcacatgacatcttcttgatctactatgaaatttcttac 480
DB      421 aggttatttgatcacaatcacatgacatcttcttgatctactatgaaatttcttac 480
QY      481 caattaaatttcgaattcatalctcttgatttaataataacagatgaaatctcc 540
DB      481 caattaaatttcgaattcatalctcttgatttaataataacagatgaaatctcc 540
QY      541 gttatcgatcactccaatcatgatataatcttggtaactcagcaaatattaca 600
DB      541 gttatcgatcactccaatcatgatataatcttggtaactcagcaaatattaca 600
QY      601 agagatttgaagaaaaacccgaaaaatagaagaagagtagtaccatgagatg 660
DB      601 agagatttgaagaaaaacccgaaaaatagaagaagagtagtaccatgagatg 660
QY      661 tgaataattatcaaaaagataagagatgacaaacaaagtgtagtaaatgagccct 720
DB      661 tgaataattatcaaaaagataagagatgacaaacaaagtgtagtaaatgagccct 720
QY      721 gccagcttctctcaacaatcaatcagaccctatcttgatcttctgatatctgataaa 780
DB      721 gccagcttctctcaacaatcaatcagaccctatcttgatcttctgatatctgataaa 780
QY      781 ttgagataacgatttgaaaaaataatttaattgtagctgataatataatgttcca 840
DB      781 ttgagataacgatttgaaaaaataatttaattgtagctgataatataatgttcca 840
QY      841 ggaattgcaaatctcttcgttaagaacatatttgcctcttcttggttctctct 900
DB      841 ggaattgcaaatctcttcgttaagaacatatttgcctcttcttggttctctct 900
QY      901 taactatataatcgcgagatataatgataaacaatgatatcacaaacaatgtctggg 960
DB      901 taactatataatcgcgagatataatgataaacaatgatatcacaaacaatgtctggg 960
QY      961 accatttgaataaattttctcaacaattcggagacactgagaccgaccttaataa 1020
DB      961 accatttgaataaattttctcaacaattcggagacactgagaccgaccttaataa 1020
QY      1021 cgaatttacagctcactagttgagattactaagacaagaacaaagagcccgcttaagc 1080
DB      1021 cgaatttacagctcactagttgagattactaagacaagaacaaagagcccgcttaagc 1080
QY      1081 tattatacaagaattcaacaactgaatagcttgaaatcccttaagaanaatttgaatta 1140
DB      1081 tattatacaagaattcaacaactgaatagcttgaaatcccttaagaanaatttgaatta 1140
QY      1141 ccggtgttatgtgaaatatagatttgggttaacaatatgttaataatgattggtca 1200
DB      1141 ccggtgttatgtgaaatatagatttgggttaacaatatgttaataatgattggtca 1200
QY      1201 acatatataatcttcttcagaaaaaacaacttaagaagtttaacatccatata 1260
DB      1201 acatatataatcttcttcagaaaaaacaacttaagaagtttaacatccatata 1260

```

QY 1261 gggatgctataccttccacgctatgctactagagactaaagaatgctatgctgc 1320
|||||
Db 1261 gggatgctataccttccacgctatgctactagagactaaagaatgctatgctgc 1320
QY 1321 gataaatgtaattccacacgctggtgtaataatgagaccgtatgtacgatactgcga 1380
|||||
Db 1321 gataaatgtaattccacacgctggtgtaataatgagaccgtatgtacgatactgcga 1380
QY 1381 atatcatcttctggttgtaacaacataaaaaaacaagaanaaaagaatgtttt 1440
|||||
Db 1381 atatcatcttctggttgtaacaacataaaaaaacaagaanaaaagaatgtttt 1440
QY 1441 ctggtatctcatcctaatagtacttaaaatgtagatcttttgggttcaagttcgaagtc 1500
|||||
Db 1441 ctggtatctcatcctaatagtacttaaaatgtagatcttttgggttcaagttcgaagtc 1500
QY 1501 tctcaacagcgttaacacatctgcacactataaattgtcttcttaatgcatcttaacat 1560
|||||
Db 1501 tctcaacagcgttaacacatctgcacactataaattgtcttcttaatgcatcttaacat 1560
QY 1561 atttatgtttagttagaatttaataagaagcgaactgttaacattataattatag 1620
|||||
Db 1561 atttatgtttagttagaatttaataagaagcgaactgttaacattataattatag 1620
QY 1621 ataactagatgtagatattccaaatacacacttggatgttlaaacttaactgtttct 1680
|||||
Db 1621 ataactagatgtagatattccaaatacacacttggatgttlaaacttaactgtttct 1680
QY 1681 tctacagctataaatttaatacgcaggtlaaaaaagtttgccttaatttcgcgagtc 1740
|||||
Db 1681 tctacagctataaatttaatacgcaggtlaaaaaagtttgccttaatttcgcgagtc 1740
QY 1741 atgaagaataaacttaacttaattttttgaaaatgtaacccttatacctatagat 1800
|||||
Db 1741 atgaagaataaacttaacttaattttttgaaaatgtaacccttatacctatagat 1800
QY 1801 aattacgctatgcttttctgtgcataatgacacgctctacaactgtgagatcaatttc 1860
|||||
Db 1801 aattacgctatgcttttctgtgcataatgacacgctctacaactgtgagatcaatttc 1860
QY 1861 tctgcaaatatlaaatttaggaattccaatgctactatcaatagaagaacagctgagatc 1920
|||||
Db 1861 tctgcaaatatlaaatttaggaattccaatgctactatcaatagaagaacagctgagatc 1920
QY 1921 acctttaatttaagaacaaaatttttggaaaatggtataatttctcaacataattata 1980
|||||
Db 1921 acctttaatttaagaacaaaatttttggaaaatggtataatttctcaacataattata 1980
QY 1981 aataatgctataatgtaatttccatgttcttaaaaatttttttttttttttttttttt 2040
|||||
Db 1981 aataatgctataatgtaatttccatgttcttaaaaatttttttttttttttttttttt 2040
QY 2041 taaatcatatgaacaaataatgattgtgtaattccaatatactccattataattttttg 2100
|||||
Db 2041 taaatcatatgaacaaataatgattgtgtaattccaatatactccattataattttttg 2100
QY 2101 aaatcccaaatattataatttttagtccaataacatgcatagaagaattccaaaaaaatt 2160
|||||
Db 2101 aaatcccaaatattataatttttagtccaataacatgcatagaagaattccaaaaaaatt 2160
QY 2161 ttgttaacgaacatcccaaattttttttttttttttttttttttttttttttttttttt 2220
|||||
Db 2161 ttgttaacgaacatcccaaattttttttttttttttttttttttttttttttttttttt 2220
QY 2221 ctattttgttggaaatggaagtagtaataatatacgaagaattttaaaaaattatata 2280
|||||
Db 2221 ctattttgttggaaatggaagtagtaataatatacgaagaattttaaaaaattatata 2280
QY 2281 agcctataagcgtcgaagaatgatttctctagtggtgtaattataatgtagtgat 2340
|||||
Db 2281 agcctataagcgtcgaagaatgatttctctagtggtgtaattataatgtagtgat 2340
QY 2341 tcagaattggacaacatgaaaaaaggaatttaaatatttaactttaaaataaataat 2400
|||||

Db 2341 tcagaattggacaacatgaaaaaaggaatttaaatatttaactttaaaataaataat 2400
|||||
QY 2401 ttgagttaaatgtgttttctgacataatgaggggcaaaaaaagaacatgccaaagtctac 2460
|||||
Db 2401 ttgagttaaatgtgttttctgacataatgaggggcaaaaaaagaacatgccaaagtctac 2460
QY 2461 gggttgacttcaggttcggaataatccaataactgtcttttgaacgcaagctcgtg 2520
|||||
Db 2461 gggttgacttcaggttcggaataatccaataactgtcttttgaacgcaagctcgtg 2520
QY 2521 taagggttccttcggaatttccactgttccacctactcgtgtagccaccctttccca 2580
|||||
Db 2521 taagggttccttcggaatttccactgttccacctactcgtgtagccaccctttccca 2580
QY 2581 tatcctaagggttaatttttggaaatcccaatttaaacggatgtagccgtaacggaactcc 2640
|||||
Db 2581 tatcctaagggttaatttttggaaatcccaatttaaacggatgtagccgtaacggaactcc 2640
QY 2641 tgggattctgctggagcattatcaaaaatttatagacgaagtgggttataattataa 2700
|||||
Db 2641 tgggattctgctggagcattatcaaaaatttatagacgaagtgggttataattataa 2700
QY 2701 aactcaacacttgatcagataaaatttcaataaacactttagcgtatcgtacgactc 2760
|||||
Db 2701 aactcaacacttgatcagataaaatttcaataaacactttagcgtatcgtacgactc 2760
QY 2761 atctaatgacttttttttctccacgctgtagtaagaatgatactattagccagag 2820
|||||
Db 2761 atctaatgacttttttttctccacgctgtagtaagaatgatactattagccagag 2820
QY 2821 acatgattatgataatataccaattcaatgatatatttagtataaagtgttaaa 2880
|||||
Db 2821 acatgattatgataatataccaattcaatgatatatttagtataaagtgtgttaa 2880
QY 2881 ctatttcagcatcgacgcttctgcgaactttgttttlaaatttaagaagtttaataaa 2940
|||||
Db 2881 ctatttcagcatcgacgcttctgcgaactttgttttlaaatttaagaagtttaataaa 2940
QY 2941 aagtaataaaggagacataagcgaacaaagaatgatacagcaggaacaaagcca 3000
|||||
Db 2941 aagtaataaaggagacataagcgaacaaagaatgatacagcaggaacaaagcca 3000
QY 3001 tgaagctcatgtgttgaattgaatttaagaagaatttttaatttaagaagatg 3060
|||||
Db 3001 tgaagctcatgtgttgaattgaatttaagaagaatttttaatttaagaagatg 3060
QY 3061 atacaattatatttctgactctttaaaccctcttaacaacaggaagctcccttt 3120
|||||
Db 3061 atacaattatatttctgactctttaaaccctcttaacaacaggaagctcccttt 3120
QY 3121 tcagtagaagtcgcgaattcccaactttaaagaacaaagcattagaagaagaatgtagtga 3180
|||||
Db 3121 tcagtagaagtcgcgaattcccaactttaaagaacaaagcattagaagaagaatgtagtga 3180
QY 3181 gagagagaggaagaactagctcc 3202
|||||
Db 3181 gagagagaggaagaactagctcc 3202

RESULT 2

AAF58252/C

ID AAF58252 standard; DNA; 936 BP.

XX AAF58252;

XX 24-APR-2001 (first entry)

DE Oligonucleotide D1835.

XX Electron-transfer group; ETM; mismatch; genotyping;

KW gene expression; ss.

XX

XX (CLIN-) CLINICAL MICRO SENSORS INC.
PA Umek RM;
XX WPI: 2001-159728/16.
XX
XX Nucleic acids containing electron-transfer group, useful as labels in
PT hybridization assays, e.g. for genotyping, allowing repeat analyses on
PT a single surface
XX
XX Example 6; Page 128; 159pp; English.
PS
XX The present invention relates to a composition comprising two nucleic
CC acids each containing an electron-transfer group (ETM) having
CC different redox potentials. The invention is used for electronic
CC detection of nucleic acids, especially of substitutions (mismatches)
CC and single-nucleotide polymorphisms, e.g. for genotyping,
CC monitoring gene expression.
XX
SQ Sequence 936 BP; 5 A; 139 C; 10 G; 6 T; 776 other;

Query Match 7.8%; Score 251; DB 22; Length 936;
Best Local Similarity 1.0%; Pred. No. 7.7e-31;
Matches 8; Conservative 589; Mismatches 184; Indels 0; Gaps 0;

OY 1508 ggcgttaaccatcgtcaacttaattgcttcttaatgcatcttaacattatt 1567
DB 793 GGGGGCWW 734
OY 1568 gttagtgaatttaagaagcgaactgttaacattataattatagatactag 1627
DB 733 WWW 674
OY 1628 tatgtatattccaatacacttgatgttaacttaattctgttctcctacg 1687
DB 673 WWW 614
OY 1688 gataataatatacatcgcgaagtaaaaaagttgtctatttccgcgatacgaag 1747
DB 613 WWW 554
OY 1748 ataaacctaatgacttaatttttgaaaatgaaccttaccatagataattacc 1807
DB 553 WWW 494
OY 1808 gtagtctttgtgcataatgacagcctctacaactgtgatacttaattttctgcaa 1867
DB 493 WWW 434
OY 1868 atattaatlagaaatcactgctactatcaatagagaagaacgcgataattacatt 1927
DB 433 WWW 374
OY 1928 aatttaagaacaaattttgaaaaatgtataatttcaacataattataaatatga 1987
DB 373 WWW 314
OY 1988 tgcctataatgattcctatgtcttaaaatatttttataattagttataaac 2047
DB 313 WWW 254
OY 2048 attagaaccataatagttgtgaattcaaatatccatataatttttgaatc 2107
DB 253 WWW 194
OY 2108 caaatatataattagtcataacatgcatagaaagttccaaaaaatttggttaa 2167
DB 193 WWW 134
OY 2168 cagaacttcaaatctttttttataggaacagaataacagatgaaactattt 2227
DB 133 WWW 74
OY 2228 gtttggaaatggaagtagcataacataagcaaattttaaaaaattataagcctat 2287
DB 73 WWW 14
OY 2288 a 2288
DB 13 W 13

DB 133 WWW 74
OY 2228 gtttggaaatggaagtagcataacataagcaaattttaaaaaattataagcctat 2287
DB 73 WWW 14
OY 2288 a 2288
DB 13 W 13

RESULT 7

AAFS825/C

ID AAF58255 strand; DNA; 938 BP.

AC AAF58255;

DT 24-APR-2001 (first entry)

DE Oligonucleotide D1876.

KW Electron-transfer group; ETM; mismatch; genotyping;

KW gene expression; ss.

OS Synthetic.

PN W0200107665-A2.

PD 01-FEB-2001.

PF 26-JUL-2000; 2000WO-US20476.

PR 26-JUL-1999; 99US-0145695.

PR 17-MAR-2000; 2000US-0190259.

XX (CLIN-) CLINICAL MICRO SENSORS INC.

XX Umek RM;

XX WPI: 2001-159728/16.

XX Nucleic acids containing electron-transfer group, useful as labels in
PT hybridization assays, e.g. for genotyping, allowing repeat analyses on
PT a single surface
XX

XX Example 6; Page 127; 159pp; English.

XX The present invention relates to a composition comprising two nucleic
CC acids each containing an electron-transfer group (ETM) having
CC different redox potentials. The invention is used for electronic
CC detection of nucleic acids, especially of substitutions (mismatches)
CC and single-nucleotide polymorphisms, e.g. for genotyping,
XX monitoring gene expression.

SQ Sequence 938 BP; 4 A; 144 C; 9 G; 5 T; 776 other;

Query Match 7.8%; Score 251; DB 22; Length 938;
Best Local Similarity 1.0%; Pred. No. 7.7e-31;
Matches 8; Conservative 589; Mismatches 184; Indels 0; Gaps 0;

OY 1508 ggcgttaaccatcgtcaacttaattgcttcttaatgcatcttaacattatt 1567
DB 793 GGGGGCWW 734
OY 1568 gttagtgaatttaagaagcgaactgttaacattataattatagatactag 1627
DB 733 WWW 674
OY 1628 tatgtatattccaatacacttgatgttaacttaattctgttctcctacg 1687
DB 673 WWW 614

OY 2324 ataat 2328
DB 781 www 785

RESULT 9

AAFS8254
ID AAF58254 standard; DNA; 936 BP.

AC AAF58254;

DT 24-APR-2001 (first entry)

DE Oligonucleotide D1875.

KW Electron-transfer group; ETM; mismatch; genotyping;
gene expression; ss.

OS Synthetic.

PN WO200107665-A2.

PD 01-FEB-2001.

PF 26-JUL-2000; 2000WO-US20476.

PR 26-JUL-1999; 99US-0145695.

PR 17-MAR-2000; 2000US-0190259.

PA (CLIN-) CLINICAL MICRO SENSORS INC.

PI Umek RM;

DR WPI: 2001-159728/16.

XX Nucleic acids containing electron-transfer group, useful as labels in
XX hybridization assays, e.g. for genotyping, allowing repeat analyses on
XX a single surface

PS Example 6; Page 127; 159pp; English.

CC The present invention relates to a composition comprising two nucleic
CC acids each containing an electron-transfer group (ETM) having
CC different redox potentials. The invention is used for electronic
CC detection of nucleic acids, especially of substitutions (mismatches)
CC and single-nucleotide polymorphisms, e.g. for genotyping,
CC monitoring gene expression.

SQ Sequence 936 BP; 4 A; 144 C; 7 G; 5 T; 776 other;

Query Match 7.8%; Score 250.6; DB 22; Length 936;

Best Local Similarity 0.5%; Pred. No. 8.8e-31;

Matches 4; Conservative 596; Mismatches 185; Indels 0; Gaps 0;

OY 1544 taatgacattacaatatattgttagtgaattataagcggaactgttaact 1603

DB 1 www 60

OY 1604 tacaatattatagatagatagatgattccaatcacatgacttgattta 1663

DB 61 www 120

OY 1664 aacttaattcttcttctcagataaataatcaatcagagtaaaaaagtttg 1723

DB 121 www 180

OY 1724 tctatttcgagcagcagagataaaccataagacttaattttgaaatgttaac 1783

DB 181 www 240

OY 1784 ccttactacatagattaatcagatgtttttgttcacataagacgcctctaacac 1843

DB 241 www 300

OY 1844 tgtatagtcatttttctcgaatataataggaattcaatgctactacataga 1903

DB 301 www 360

OY 1904 agaaacgctggtatctacatttaattaaagacaaattttgaaaaatgttaact 1963

DB 361 www 420

OY 1964 tctacaatatattaaaatagatgcataatgtattctcctagttcttaaatatt 2023

DB 421 www 480

OY 2024 tttttatatttagtataataacatttagaaccaataagttggtgaattcaatlc 2083

DB 481 www 540

OY 2084 tccattaatattttgaaatctacaattataatttagtcaataacatgatataga 2143

DB 541 www 600

OY 2144 aagttccaaaaaatttgttaacagaaactccaatttttttttaigaaacag 2203

DB 601 www 660

OY 2204 aataacagatagaaactatttgttggaatggaatgataatataacaaagcaaa 2263

DB 661 www 720

OY 2264 tttaaataattataagcctacacgcctcaagtagtatctagtagtgaatta 2323

DB 721 www 780

OY 2324 ataat 2328

DB 781 www 785

RESULT 10

AAFS8257

ID AAF58257 standard; DNA; 936 BP.

AC AAF58257;

DT 24-APR-2001 (first entry)

DE Oligonucleotide D1954.

KW Electron-transfer group; ETM; mismatch; genotyping;
gene expression; ss.

OS Synthetic.

PN WO200107665-A2.

PD 01-FEB-2001.

PF 26-JUL-2000; 2000WO-US20476.

PR 26-JUL-1999; 99US-0145695.

PR 17-MAR-2000; 2000US-0190259.

PA (CLIN-) CLINICAL MICRO SENSORS INC.

PI Umek RM;

DR WPI: 2001-159728/16.

XX Nucleic acids containing electron-transfer group, useful as labels in
XX hybridization assays, e.g. for genotyping, allowing repeat analyses on
XX a single surface


```

Db 301 wwwwwwwww 360
Qy 1904 agaaacagtgatgtacatttaatttaagaacaaattttgaaatgtataatt 1963
Db 361 wwwwwwwww 420
Qy 1964 tctaacaatatttaaaatagatgcctataatgatttcctatgttcttaaatatt 2023
Db 421 wwwwwwwww 480
Qy 2024 tttttatattagtataataacattagaaacaaatagttggtgaattcaatc 2083
Db 481 wwwwwwwww 540
Qy 2084 tccattaatatttttgaatctacaattatattatagtcataacaaatgcata 2143
Db 541 wwwwwwwww 600
Qy 2144 aagttccaaaaaaatttgttaacagaaactccaattttttttatggaacaag 2203
Db 601 wwwwwwwww 660
Qy 2204 aaataacagatgaaactatttltgtggaatggaagtagtaataacaaatga 2263
Db 661 wwwwwwwww 720
Qy 2264 ttttaaaaaattataagcctatacgcgctcaaatgatttctcagtagtgtaata 2323
Db 721 wwwwwwwww 780
Qy 2324 ataatt 2328
Db 781 wwwwww 785

RESULT 12
AAFS8262
ID AAF58262 standard; DNA; 936 BP.
XX
AC AAF58262;
XX
DT 24-APR-2001 (first entry)
XX
DE Oligonucleotide D2007.
XX
KW Electron-transfer group; ETM; mismatch; genotyping;
XX gene expression; ss.
XX
OS Synthetic.
XX
PN WO200107665-A2.
XX
PD 01-FEB-2001.
XX
PF 26-JUL-2000; 2000WO-US20476.
XX
PR 26-JUL-1999; 99US-0145695.
XX 17-MAR-2000; 2000US-0190259.
XX
PA (CLIN-) CLINICAL MICRO SENSORS INC.
XX
PI Umek RM;
XX
DR WPI; 2001-159728/16.
XX
PT Nucleic acids containing electron-transfer group, useful as labels in
XX hybridization assays, e.g. for genotyping, allowing repeat analyses on
XX a single surface
XX
PS Example 6; Page 128; 159pp; English.
XX
CC The present invention relates to a composition comprising two nucleic
XX acids each containing an electron-transfer group (ETM) having

```

```

CC different redox potentials. The invention is used for electronic
CC detection of nucleic acids, especially of substitutions (mismatches)
CC and single-nucleotide polymorphisms, e.g. for genotyping,
CC monitoring gene expression.
XX
SQ Sequence 936 BP; 5 A; 139 C; 10 G; 6 T; 776 other;

Query Match      7 8%; Score 250.6; DB 22; Length 936;
Best Local Similarity 0.58; Pred. No. 8,8e-31;
Matches 4; Conservative 596; Mismatches 185; Indels 0; Gaps 0;

Qy 1544 taatgcatcttaacattattgttagttggaatttaagaagcgaactgttaact 1603
Db 1 wwwwwwwww 60
Qy 1604 tacaataattataatagataactagtagtattccaataacatacttgagattta 1663
Db 61 wwwwwwwww 120
Qy 1664 aacttaactgtttcttctcctaagcataaattatcatcagagtgtaaaaaagtttg 1723
Db 121 wwwwwwwww 180
Qy 1724 tctatttcgcgtagtgaagataaacctaactgaacttaatttttgaaaatgtaac 1783
Db 181 wwwwwwwww 240
Qy 1784 cctttactcatagatttaacttaacgtagtlttltgtgcataatgacgctctaac 1843
Db 241 wwwwwwwww 300
Qy 1844 tgtgatagtcataattttctgcgaatataatgaattgaattcactatcactaga 1903
Db 301 wwwwwwwww 360
Qy 1904 agaaacagctgattacatttaatttaagaacaaattttgaaatgtataatt 1963
Db 361 wwwwwwwww 420
Qy 1964 tctaacaatatttaaaatagatgcctataatgatttcctatgttcttaaatatt 2023
Db 421 wwwwwwwww 480
Qy 2024 tttttatattagtataataacattagaaacaaatagttggtgaattcaatc 2083
Db 481 wwwwwwwww 540
Qy 2084 tccattaatatttttgaatctacaattatattatagtcataacaaatgcata 2143
Db 541 wwwwwwwww 600
Qy 2144 aagttccaaaaaaatttgttaacagaaactccaattttttttatggaacaag 2203
Db 601 wwwwwwwww 660
Qy 2204 aaataacagatgaaactatttltgtggaatggaagtagtaataacaaatga 2263
Db 661 wwwwwwwww 720
Qy 2264 ttttaaaaaattataagcctatacgcgctcaaatgatttctcagtagtgtaata 2323
Db 721 wwwwwwwww 780
Qy 2324 ataatt 2328
Db 781 wwwwww 785

RESULT 13
AAFS8255
ID AAF58255 standard; DNA; 938 BP.
XX

```


[illegible]

Search completed: October 17, 2001, 19:18:28
Job time: 15908 sec

sequence 244 BP; 19 A; 9 C; 12 G; 10 T; 194 other;

Query Match 2.7%; Score 85.4; DB 22; Length 244;
Best Local Similarity 3.9%; Pred. No. 3.1e-05;
Matches 8; Conservative 163; Mismatches 34; Indels 0; Gaps 0

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 17, 2001, 14:52:05 ; Search time 156 Seconds

(without alignments)
3885.732 Million cell updates/sec

Title: US-09-502-426-1_COPY_1_3202

Perfect score: 3202

Sequence: 1 atgtggtatatatatgttg.....gagagagagaactagctcc 3202

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 324599 seqs, 94655562 residues

Total number of hits satisfying chosen parameters: 649198

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database : Issued_Patents_NA:*
1: /cgn2_6/ptodata/2/ina/5A.COMB.seq:*
2: /cgn2_6/ptodata/2/ina/5B.COMB.seq:*
3: /cgn2_6/ptodata/2/ina/6A.COMB.seq:*
4: /cgn2_6/ptodata/2/ina/6B.COMB.seq:*
5: /cgn2_6/ptodata/2/ina/PTCUS.COMB.seq:*
6: /cgn2_6/ptodata/2/ina/backfile1.seq:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description |
|------------|-------|-------------|--------|----|--------------------|
| 1 | 77.4 | 2.4 | 19124 | 2 | US-08-487-826B-13 |
| 2 | 68.6 | 2.1 | 837 | 4 | US-08-998-416-288 |
| 3 | 67.4 | 2.1 | 19124 | 2 | US-08-487-826B-13 |
| 4 | 65.8 | 2.1 | 615 | 4 | US-08-998-416-186 |
| 5 | 65.4 | 2.0 | 636 | 4 | US-08-998-416-1137 |
| 6 | 63.6 | 2.0 | 8920 | 2 | US-08-446-855A-1 |
| 7 | 63.6 | 2.0 | 8920 | 4 | US-09-150-741-1 |
| 8 | 62.6 | 2.0 | 4818 | 3 | US-08-817-926-27 |
| 9 | 61.8 | 1.9 | 5852 | 1 | US-07-867-106-2 |
| 10 | 60.8 | 1.9 | 4818 | 3 | US-08-817-926-27 |
| 11 | 60.6 | 1.9 | 837 | 4 | US-08-998-416-288 |
| 12 | 60.4 | 1.9 | 8920 | 2 | US-08-446-855A-1 |
| 13 | 60.4 | 1.9 | 8920 | 4 | US-09-150-741-1 |
| 14 | 60.2 | 1.9 | 665 | 2 | US-08-883-795A-36 |
| 15 | 59.2 | 1.8 | 51952 | 3 | US-08-947-823-1 |
| 16 | 59 | 1.8 | 827 | 4 | US-08-998-416-535 |
| 17 | 58.4 | 1.8 | 665 | 2 | US-08-883-795A-36 |
| 18 | 57.8 | 1.8 | 1511 | 1 | US-07-991-867B-8 |
| 19 | 57.8 | 1.8 | 1511 | 1 | US-08-107-755A-8 |
| 20 | 57.8 | 1.8 | 1511 | 1 | US-08-544-332-8 |
| 21 | 57.4 | 1.8 | 3095 | 6 | 5231168-1 |
| 22 | 56.4 | 1.8 | 2058 | 3 | US-08-749-391-1 |
| 23 | 56.4 | 1.8 | 2058 | 3 | US-09-390-200-1 |
| 24 | 56.4 | 1.8 | 4673 | 1 | US-07-638-431-1 |
| 25 | 56.4 | 1.8 | 4673 | 5 | PCT-US92-00018-1 |
| 26 | 56.2 | 1.8 | 2960 | 3 | US-08-913-842-3 |
| 27 | 55.6 | 1.7 | 701 | 4 | US-08-998-416-701 |

| | | | | | | |
|----|------|-----|------|---|-------------------|-------------------|
| 28 | 55.6 | 1.7 | 5852 | 1 | US-07-867-106-2 | Sequence 2, Appl |
| 29 | 55.2 | 1.7 | 1406 | 3 | US-08-913-842-6 | Sequence 6, Appl |
| 30 | 55.2 | 1.7 | 1511 | 1 | US-07-991-867B-8 | Sequence 8, Appl |
| 31 | 55.2 | 1.7 | 1511 | 1 | US-08-107-755A-8 | Sequence 8, Appl |
| 32 | 55.2 | 1.7 | 1511 | 2 | US-08-544-332-8 | Sequence 8, Appl |
| 33 | 55 | 1.7 | 1850 | 3 | US-08-617-860B-32 | Sequence 32, Appl |
| 34 | 55 | 1.7 | 4098 | 2 | US-08-605-106-4 | Sequence 4, Appl |
| 35 | 54.8 | 1.7 | 6768 | 1 | US-08-107-755A-1 | Sequence 1, Appl |
| 36 | 54.8 | 1.7 | 8457 | 1 | US-07-991-867B-1 | Sequence 1, Appl |
| 37 | 54.8 | 1.7 | 8457 | 2 | US-08-544-332-1 | Sequence 1, Appl |
| 38 | 54.6 | 1.7 | 9048 | 3 | US-08-973-273-4 | Sequence 4, Appl |
| 39 | 54.2 | 1.7 | 663 | 4 | US-08-998-416-191 | Sequence 191, App |
| 40 | 53.8 | 1.7 | 6243 | 2 | US-09-056-075-1 | Sequence 1, Appl |
| 41 | 53.6 | 1.7 | 2430 | 4 | US-08-845-258-3 | Sequence 3, Appl |
| 42 | 53.6 | 1.7 | 2430 | 4 | US-08-845-258-40 | Sequence 40, Appl |
| 43 | 53.6 | 1.7 | 2430 | 4 | US-08-990-571-3 | Sequence 3, Appl |
| 44 | 53.6 | 1.7 | 2430 | 4 | US-08-990-571-40 | Sequence 40, Appl |
| 45 | 53.2 | 1.7 | 1850 | 3 | US-08-617-860B-32 | Sequence 32, Appl |

ALIGNMENTS

RESULT 1
US-08-487-826B-13
; Sequence 13, Application US/08487826B
; Patent No. 5993827
; GENERAL INFORMATION:
; APPLICANT: Sim, Kim L.
; APPLICANT: Chitnis, Chetan
; APPLICANT: Miller, Louis H.
; APPLICANT: Peterson, David S.
; APPLICANT: Su, Xin-zhuan
; APPLICANT: Wellens, Thomas E.
; TITLE OF INVENTION: BINDING DOMAINS FROM PLASMODIUM VIVAX
; TITLE OF INVENTION: AND PLASMODIUM FALCIPARUM ERYTHROCYTE BINDING PROTEINS
; NUMBER OF SEQUENCES: 45
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Knobbe Martens Olson & Bear
; STREET: 620 Newport Center Drive 16th Floor
; CITY: Newport Beach
; STATE: California
; COUNTRY: US
; ZIP: 92660
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/487,826B
; FILING DATE: 10-SEP-1993
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Israelson, Ned
; REGISTRATION NUMBER: 29,655
; REFERENCE/DOCKET NUMBER: NIH121.001CPI
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 235-8550
; TELEFAX: (619) 235-0176
; INFORMATION FOR SEQ ID NO: 13:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 19124 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; US-08-487-826B-13
Query Match 2.4%; Score 77.4; DB 2; Length 19124;

Best Local Similarity 42.6%; Pred. No. 1.2e-06;
Matches 590; Conservative 0; Mismatches 781; Indels 14; Gaps 3;

```

OY 1013 ttaaaataagattcaacgctcactagctgagatcactagataagaagaagacc 1072
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 229 TTATATATACAAATATATATATAGTTTCCCTATTAATAATTAATATATATATATAT 288
OY 1073 gtccaagctattatacaaaagttacaactgaaatagcttgaactccttgaagaatt 1132
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 289 TTTCTGTATTTTATTAATAATATACAAATTTCTTATTTTATTTTAACTTATTCCTTTT 348
OY 1133 tggaaattaccggtgtgtatgataatagattagtgtgtaacaaataagtaacaaatt 1192
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 349 AATTTCTTAATTTCTTTATGCAACAAAAACATAAGTAATTCATCAATATCAACAAAAA 408
OY 1193 agtggcaacatacatataactccttaacagaaaacaactagaagaagtaacatat 1252
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 409 AAAAAAAAAAAAAAAAAAAAAAAAAATTTATATATATATATATATATATATATATATAT 468
OY 1253 ccataatagtgatagctatcacttcaacgtatgctatcactagaagaataagatgat 1312
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 469 TCACCTATATATATATATATATATATATATATATATATATATATATATATATATAT 528
OY 1313 gtgatgctgataaagaaattcaacgctgtgtaataattatgggaacgctatcgtacgat 1372
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 529 TTTTAACATAGAAAGATTAGAAATACATTTTCTTTTCTTGAATGCTAATTCACACAT 588
OY 1373 cactgcaaatatcattcctgtgtgtgtcaacaaataaaacaagaagaagaaga 1432
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 589 ATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 648
OY 1433 cgaatttcttgatccatcactcaatgatacgaatgataccttcttgatgataagtt 1492
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 649 TCTTTTATTC-----ACATGTGAATAATATATATATATATATATATATATATAT 702
OY 1493 cgaagtcctcacaagcggtgatacactcgcgaactaataattgcttcttcaagcatt 1552
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 703 ATGCTGTTATATATATATATATATATATATATATATATATATATATATATATATAT 762
OY 1553 ttaacaataattcgttgaattcgaatttaagaagcgaacttgaacttacaatat 1612
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 763 AATATATATATATATATATATATATATATATATATATATATATATATATATATAT 822
OY 1613 tatattagatagatgatagtatattccaaatacacttggatgtttaaactaatc 1672
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 823 TTTATTTTCAAAATACATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 882
OY 1673 ttgttcttccctcaggtataaataataatcactcgaaggtaaaagtttgcctaat 1732
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 883 GCATATACATGCACATTTTAAATTAATTAATAAATTTTATTTTAAATATATATATACA 942
OY 1733 cgcgaatgcaagaagaataaccataagcttcaatttttgaagaatgtaacccttcaat 1792
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 943 ATTTTCATATACATTAATTTTTCACACACATTTAAGTGTCTATATATATATATATATAT 1002
OY 1793 catagattaaattaccgtaagttttgtgcataaagcgaacttacaactgtatagt 1852
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1003 AATATATATATATATATATATATATATATATATATATATATATATATATATATAT 1062
OY 1853 caatttttctgcgaataataaataaggaattcaatgctacatacagaagaagcgc 1912
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1063 CTGTATTTATCATATATATATATATATATATATATATATATATATATATATATATA 1122
OY 1913 tgaagattacatttcaatttgaagaagaataatttgaagaatgtaatttcaacaaat 1972
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1123 TTAATAAATATATATATATATATATATATATATATATATATATATATATATATATATA 1182
OY 1973 attataaataatgataagctataatgataattcctcattgttcttaaaataatttttata 2032
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1183 TAAGATATAAATTCATATATATATATATATATATATATATATATATATATATATATA 1237
OY 2033 tttagttataataatagataagaacaaataagttgtggaattcaacatactcattaat 2092
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

```

DB 1238 TATATATATATATATATATATATATATATATATATATATATATATATATATATAT 1294
OY 2093 atttttgaactcacaattataatttagtcaataacatgcatagaagaattccaa 2152
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1295 TTTGTATATACATACAGACTTAAGAAACTATATATATATATATATATATATATATAT 1354
OY 2153 aaaaaatttgttcaacgaactccaaatttttttttataggaaagaagaataacag 2212
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1355 ATAAATATCTTTTATATATATATATATATATATATATATATATATATATATATATAT 1414
OY 2213 atagaataactatttgtgtgtggaatggaatgataataacttaagcaatttcaaaa 2272
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1415 TATATATATTTTTCATATATATATATATATATATATATATATATATATATATATAT 1474
OY 2273 attataaagcctatacgcgcacaagatgctatcctagtagtgaattataatgcat 2332
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1475 AAAACTTCAAACATTTTTCATTAATATATATATATATATATATATATATATATATAT 1534
OY 2333 ggtgcgattcagaattggaacacatgaaacggaatataataatlaactttaaataa 2392
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1535 ACAGAGAAACGTAGAACATACCAAAAAATTAAGAACAAAAAGATATTAACAAAAATATA 1594
OY 2393 ataaa 2397
    |||
DB 1595 TAAAA 1599
    |||

RESULT 2
US-08-998-416-288
; Sequence 288, Application US/08998416
; Patent No. 6239264
; GENERAL INFORMATION:
; APPLICANT: Philippsen, Peter
; APPLICANT: Pohlmann, Rainer
; APPLICANT: Stelner, Sabine
; APPLICANT: Mohr, Christine
; APPLICANT: Wendland, Jürgen
; APPLICANT: Knechtle, Philipp
; APPLICANT: Redischung, Corinne
; TITLE OF INVENTION: GENOMIC DNA SEQUENCES OF ASHBYA GOSSYPII
; NUMBER OF SEQUENCES: 1152
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: No. 6239264artis Corporation
; STREET: 3054 Cornwalls Road
; CITY: Research Triangle Park
; STATE: No. 6239264th Carolina
; COUNTRY: USA
; ZIP: 27709
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/998,416
; FILING DATE: 24-DEC-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: CH 0016/97
; FILING DATE: 31-DEC-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Meigs, J. Timothy
; REGISTRATION NUMBER: 38,241
; REFERENCE/DOCKET NUMBER: PF/5-30306/A/CGC1976
; TELEPHONE: 919-541-8587
; TELEFAX: 919-541-8689
; INFORMATION FOR SEQ ID NO: 288:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 837 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single

```

```

;      TOPOLOGY: linear
;      MOLECULE TYPE: DNA (genomic)
;      ORIGINAL SOURCE:
;      ORGANISM: PAG1241RP
US-08-998-416-288

```

| | | | | | |
|---------------------------|-------|-----------------|------|------------|--------|
| Query Match | 2.1% | Score 68.6 | DB 4 | Length 837 | |
| Best Local Similarity | 46.9% | Pred. No. 3e-05 | | | |
| Matches 250, Conservative | 0 | Mismatches 279 | | Indels 4 | Gaps 1 |

| | | | |
|----|------|--|------|
| OY | 1750 | aaacctaatagacttaatttttggaaaagtaacccttttaetcatgaatcaattaccgt | 1809 |
| Db | 56 | AATATTAGTAAATAAATAATTAACCTATTATTATTCATATTATTAATAATTATTTG | 115 |
| OY | 1810 | atgcttgtagcacaataagcacgcctccacaacigtgtagtaatttttcgcaaat | 1869 |
| Db | 116 | ATTATTTAATCTATRTATATAAATTANTATATATATTACTTAATTCATATTTAATYT | 175 |
| OY | 1870 | attaataatgaatticaatgctactactcaataagaagaacagctagattacatttaa | 1929 |
| Db | 176 | TATATATTATAAAAATAAATATTATATATGAATACATTTACTCATGTCCAATTTTA | 235 |
| OY | 1930 | tlttaagacaaaatlftggaanaagtctaaatlctcaacaatatlaaanaatatgtg | 1989 |
| Db | 236 | ATTGGTATTAAAAATATTATTAGATATTTATTTCTTTATTAATAATTATTAATAGATY | 295 |
| OY | 1990 | cctaatagtattlccatagcttgtcttaaaatatlttt---tlttatattgattataat | 2045 |
| Db | 296 | ATCAATATATTATATATTATTATTAATTAAGTGTTTTAAAAATATATTTATTTATTTATA | 355 |
| OY | 2046 | acattatgaaccaataatagttgtagaattccaataatcccataatattttgaaac | 2105 |
| Db | 356 | AAGATTATTAATTATTTAAATATGTAAATATTTATTTATTTATATAATATCATTTTTTATA | 415 |
| OY | 2106 | tacaataathataatatttagtccaataacaatgatagaagaagttccaanaaaaaattfgt | 2165 |
| Db | 416 | AATATTATAGTGAATTTATTTATTTATTAATCGTTTATTAAGAATTATTTATTAATAATYT | 475 |
| OY | 2166 | aacggaaacctccaaatlftttttttagtgacaagaataacaagatagaaaaactat | 2225 |
| Db | 476 | TAACTTAAATTTCTTATTATTATTTAAATTTTATATTTAATAAATTATATTTCCATTTTATY | 535 |
| OY | 2226 | tltgttggaagtggaagtagtaataacatlaagcaaatltttaaaaaattata | 2278 |
| Db | 536 | TATTTATTTATTTAATTAATTAATTTAATTTAATTAATTTATTAATTCATTAATTTTA | 588 |

RESULT 3
 US-08-487-826B-13/c
 : Sequence 13, Application US/08487826B
 : Patent No. 5993827
 :
 GENERAL INFORMATION:
 :
 APPLICANT: Sim, Kim L.
 :
 APPLICANT: Chitnis, Chetan
 :
 APPLICANT: Miller, Louis H.
 :
 APPLICANT: Peterson, David S.
 :
 APPLICANT: Su, Xin-zhaun
 :
 APPLICANT: Wellens, Thomas E.
 :
 TITLE OF INVENTION: BINDING DOMAINS FROM PLASMODIUM VIVAX
 :
 TITLE OF INVENTION: AND PLASMODIUM FALCIPARUM ERYTHROCYTE BINDING PROTEINS
 :
 NUMBER OF SEQUENCES: 45
 :
 CORRESPONDENCE ADDRESSES:
 :
 ADDRESSEE: Knobbe Martens Olson & Bear
 :
 STREET: 620 Newport Center Drive 16th Floor
 :
 CITY: Newport Beach
 :
 STATE: California
 :
 COUNTRY: US
 :
 ZIP: 92660
 :
 COMPUTER READABLE FORM:
 :
 MEDIUM TYPE: Floppy disk
 :
 COMPUTER: IBM PC compatible
 :
 OPERATING SYSTEM: PC-DOS/MS-DOS
 :

```

? SOFTWARE: patentin Release #1.0, Version #1.25
?
? CURRENT APPLICATION NUMBER: US/08/487,826B
? APPLICATION NUMBER: 08/487,826
? FILING DATE: 10-SEP-1993
? CLASSIFICATION: 435
?
? ATTORNEY/AGENT INFORMATION:
? NAME: Israelson, Ned
? REGISTRATION NUMBER: 29,655
?
? REFERENCE/DOCKET NUMBER: NH121.001CP1
? TELECOMMUNICATION INFORMATION:
? TELEPHONE: (619) 235-8550
? TELEFAX: (619) 235-0176
? INFORMATION FOR SEQ ID NO: 13:
? SEQUENCE CHARACTERISTICS:
? LENGTH: 19124 base pairs
? TYPE: nucleic acid
? STRANDEDNESS: single
? TOPOLOGY: linear
? MOLECULE TYPE: cDNA
? HYPOTHETICAL: NO
? ANTI-SENSE: NO
?
US-08-487-826B-13

```

[illegible]

```

Db 6387 TATTTTATGATTTTCATGCAATTTATGCAATTTTCAAAATTTTATGATATATATATATAA 6328
Qy 1137 attaccgggttgatgtaataatagatgtagtggttaacaataatgtaataatgtagt 1196
Db 6327 AATAAGTAAATAATACACATTTTAAATATATATATATATATATATATATATATATATAT 6268
Qy 1197 gtaacataataataatccttaacagaaataacaaacttaagaagaagttaacatacat 1256
Db 6267 TTTATTTTCTATATATATTTTATATATATATATATATATATATATATATATATATAT 6208
Qy 1257 atagggtaatgataccttcac-9tatgctatactagaagactaagaagtagttagt 1315
Db 6207 TACTATATATAATATATATATATATATATATATATATATATATATATATATATATAT 6148
Qy 1316 atgctgataataagaatcacaacggtgtaataataataatgagcgtatgtaacgac 1375
Db 6147 ATMAAAATAGTATMAAAATCATCATATATATATATATATATATATATATATATATAT 6088
Qy 1376 tgaataatcatctctggtggtcaacaataaacaacaaacagaacaaagaacga 1435
Db 6087 TATATATATATATGCAATATATATGCAATATATATATATATATATATATATATATAT 6028
Qy 1436 tttctctggtatcctcaatgataatgataatgtagatgtagttagttagttagttagt 1495
Db 6027 TTTATATTTGTTATATATATATATATATATATATATATATATATATATATATATAT 5968
Qy 1496 agtccctcaacggtgtagtaacacatg--caactataatgctctccttaacgacat 1553
Db 5967 ATMAATATATAAGGATGATATATATATATATATATATATATATATATATATATATAT 5908
Qy 1554 ttaacataatctgtagttagttagttagttagttagttagttagttagttagttagttagt 1613
Db 5907 TATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 5848
Qy 1614 acattagatactagtagttagttagttagttagttagttagttagttagttagttagt 1673
Db 5847 ACTTATCTATTTTATATATATATATATATATATATATATATATATATATATATAT 5788
Qy 1674 tgttctcctcagataataataataataataataataataataataataataataata 1733
Db 5787 GTAATGCAATATCTATCTTATACACAGCATATATACCAATATATATATATATATAT 5728
Qy 1734 gtagatgataaggaataacactaagacttaatttttgaataatgtaacacttact 1793
Db 5727 TCTATGCA-CATATATAATATATATATATATATATATATATATATATATATATAT 5669
Qy 1794 atagataatcagtagttagttagttagttagttagttagttagttagttagttagt 1853
Db 5668 CCAATCAAAACACATATAATATATATATATATATATATATATATATATATATATAT 5609
Qy 1854 aattttctcgaataataataataataataataataataataataataataataata 1913
Db 5608 TATTTTGTATGAATAATCTATTTTATATATATATATATATATATATATATATATAT 5557
Qy 1914 gtagatataatlaatttaagaacaaatltttgaaatgtaataatlttaacata 1973
Db 5558 ---TATTTAAATATTTTCAAAAACATATCATGTAAATATATATATATATATATAT 5502
Qy 1974 ttattaaataatgtagcctataatgtagttagttagttagttagttagttagttagt 2033
Db 5501 TGACATATAAAATATATATATATATATATATATATATATATATATATATATATAT 5442
Qy 2034 ttagtataataatcattgaagcaataatagttgtagttagttagttagttagttagt 2090
Db 5441 TTTCTATATATACATGTTACGTAATTTGTAATTTATGTTATATATATATATATATAT 5382
Qy 2091 ---ataattttgaaatcacaataatataataatagtaataataataatgtagttagt 2147
Db 5381 GACTTATATATACGATATCTTATTTATATATATATATATATATATATATATATATAT 5322
Qy 2148 tcaaaaaaaatlttgtagcagaacatccaaatlttttttttaatgtagcagaagaat 2207

```

```

Db 5321 TATTACTTATTACTGTTGTAAGATATATATATATATATATATATATATATATATAT 5262
Qy 2208 aacagataaataactatlttgtagttagttagttagttagttagttagttagttagt 2267
Db 5261 TAAGCTATTGGAACATCAAAATCAGCATATTTATTTTCAATTTTCAATTTTCAATTA 5202
Qy 2268 aaaaaatataaagcctataagcgtcaagtagttagttagttagttagttagttagttag 2327
Db 5201 AATTATATATATATATATATATATATATATATATATATATATATATATATATATAT 5142
Qy 2328 tgcattgtagttagttagttagttagttagttagttagttagttagttagttagt 2348
Db 5141 AAAAAACACAAATATATATATATATATATATATATATATATATATATATATATAT 5121

```

```

RESULT 4
US-08-998-416-186
; Sequence 186, Application US/08998416
; Patent No. 6239264
; GENERAL INFORMATION:
; APPLICANT: Philippsen, Peter
; APPLICANT: Pohlmann, Rainer
; APPLICANT: Steiner, Sabine
; APPLICANT: Mohr, Christine
; APPLICANT: Wendland, Jurgen
; APPLICANT: Knechtle, Philipp
; APPLICANT: Reibschung, Corinne
; TITLE OF INVENTION: GENOMIC DNA SEQUENCES OF ASHBYA GOSYPII
; NUMBER OF SEQUENCES: 1152
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: No. 6239264artis Corporation
; STREET: 3054 Cornwallis Road
; CITY: Research Triangle Park
; STATE: No. 6239264th Carolina
; COUNTRY: USA
; ZIP: 27709
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/998,416
; FILING DATE: 24-DEC-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: CH 0016/97
; FILING DATE: 31-DEC-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Meigs, J. Timothy
; REGISTRATION NUMBER: 38,241
; REFERENCE/DOCKET NUMBER: PF/5-30306/A/CGC1976
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 919-541-8587
; TELEFAX: 919-541-8689
; INFORMATION FOR SEQ ID NO: 186:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 615 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: Single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; ORIGINAL SOURCE:
; ORGANISM: PAG1074RP
; US-08-998-416-186

```

```

Query Match 2.1%; Score 65.8; DB 4; Length 615;
Best Local Similarity 48.3%; Pred. No. 9.9e-05;
Matches 216; Conservative 0; Mismatches 227; Indels 4; Gaps 1;
Qy 1750 aaacctaagacttaatttttgaataatgtaacacttactacatagattactacgt 1809

```



```
COUNTRY: USA
ZIP: 22201-4714
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.24
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/446,855A
FILING DATE: 06-Jul-1995
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Mitchard, Leonard C
REGISTRATION NUMBER: 29,009
REFERENCE/DOCKET NUMBER: 47-80
TELECOMMUNICATION INFORMATION:
TELEPHONE: 703-816-4000
TELEFAX: 703-816-4100
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 8920 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: genomic
US-08-446-855A-1
```

```
Query Match 2.0%; Score 63.6; DB 2; Length 8920;
Best Local Similarity 46.8%; Pred. No. 0.00049;
Matches 278; Conservative 0; Mismatches 304; Indels 12; Gaps 2;

QY 1874 aattggaatcaatgctctatcaataagaagaacagctgattatcatattta 1933
DB 152 aaaaaataataataaaaaataaaaaataaaaaataaaaaataaaaaataa 211
QY 1934 aagacaaaatttggaaaagtata-attctaacaatatttaaaatagtcct 1992
DB 212 aatataatgattatcatatataataacatgataaaaaactttttttttct 271
QY 1993 ataagattctctatgctctaaatattttttatattagttataatcatat 2052
DB 272 ttaattatttataacattacatttgaattttttttttttttttttttttta 331
QY 2053 gaaccaataatagttggtgaattcaataatccatcaatatttttgaactacaat 2112
DB 332 tatataataataatagttggtgctgctcatttttttttttttttttttttt 391
QY 2113 tattaatattagtcataatacaatgcatagaaagttccaaaaaaatttggtaacaga 2172
DB 392 tattaataattccaatataatgataatgataatttttttttttttttttttt 451
QY 2173 acttccaattttttttttttttga-----acaagaataacagatagaaac 2221
DB 452 aatattgatttatacatttcttttttttttttttttttttttttttttttttt 511
QY 2222 tatttggttggaatggaagtatgataatcaatgaagaattttaaaaaattataa 2281
DB 512 tcccaaatatttggcttccataatttttttttttttttttttttttttttttt 571
QY 2282 gactacagcgtcgaagtatgataatgtagtgaattcaataatgcatgctgcatt 2341
DB 572 atttatttttttttttttttttttttttttttttttttttttttttttttttt 631
QY 2342 cagaattgggaacaacaatgaaacggaattaaataatcaactttaataataaataat 2401
DB 632 aaaaaaaaaaaaaaataaaaaaattttacatatgaaaaatgaaactgttatgctaaa 691
QY 2402 tgaagtaaatgttttctgactatggagggaacaaaaaagaacatgccaagaag 2455
DB 692 ttataataatttttttttttttttttttttttttttttttttttttttttttt 745
```

```
RESULT 7
US-09-150-741-1
Sequence 1, Application US/09150741
Patent No. 6183996
GENERAL INFORMATION:
APPLICANT: Stewart et al.
TITLE OF INVENTION: Nucleotide Sequence Encoding Carbamoyl Phosphate
Patent No. 6183996
TITLE OF INVENTION: Synthetase II
FILE REFERENCE:
CURRENT APPLICATION NUMBER: US/09/150,741
CURRENT FILING DATE: 1998-09-10
EARLIER APPLICATION NUMBER: PL6380
EARLIER FILING DATE: 1992-12-16
EARLIER APPLICATION NUMBER: AU93/00617
EARLIER FILING DATE: 1993-12-02
EARLIER APPLICATION NUMBER: 08/446,855
NUMBER OF SEQ ID NOS: 15
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 1
LENGTH: 8920
TYPE: DNA
ORGANISM: Plasmodium falciparum
US-09-150-741-1
```

```
Query Match 2.0%; Score 63.6; DB 4; Length 8920;
Best Local Similarity 46.8%; Pred. No. 0.00049;
Matches 278; Conservative 0; Mismatches 304; Indels 12; Gaps 2;

QY 1874 aattggaatcaatgctctatcaataagaagaacagctgattatcatattta 1933
DB 152 aaaaaataataataaaaaataaaaaataaaaaataaaaaataaaaaataa 211
QY 1934 aagacaaaatttggaaaagtata-attctaacaatatttaaaatagtcct 1992
DB 212 aatataatgattatcatatataataacatgataaaaaactttttttttct 271
QY 1993 ataagattctctatgctctaaatattttttatattagttataatcatat 2052
DB 272 ttataattataacatacatattgaatttttttttttttttttttttttttttt 331
QY 2053 gaaccaataatagttggtgaattcaataatccatcaatatttttgaactacaat 2112
DB 332 tatataataataatagttggtgctgctcatttttttttttttttttttttt 391
QY 2113 tattaatattagtcataatacaatgcatagaaagttccaaaaaaatttggtaacaga 2172
DB 392 tattaataattccaatataatgataatgataatttttttttttttttttttt 451
QY 2173 acttccaattttttttttttttga-----acaagaataacagatagaaac 2221
DB 452 atatagttttaacattcttataataaacaacatccataatattatacacatataatac 511
QY 2222 tatttggttggaatggaagtatgataatcaatgaagaattttaaaaaattataa 2281
DB 512 tcccaaatatttggcttccataatttttttttttttttttttttttttttttt 571
QY 2282 gactacagcgtcgaagtatgataatgtagtgaattcaataatgcatgctgcatt 2341
DB 572 atttatttttttttttttttttttttttttttttttttttttttttttttttt 631
QY 2342 cagaattgggaacaacaatgaaacggaattaaataatcaactttaataataaataat 2401
DB 632 aaaaaaaaaaaaaaataaaaaaattttacatatgaaaaatgaaactgttatgctaaa 691
QY 2402 tgaagtaaatgttttctgactatggagggaacaaaaaagaacatgccaagaag 2455
DB 692 ttataataatttttttttttttttttttttttttttttttttttttttttttt 745
```

RESULT 8

```

US-08-817-926-27
: Sequence 27, Application US/08817926
: Patent No. 6001590
: GENERAL INFORMATION:
: APPLICANT: Komeda, Toshihiro
: APPLICANT: Suda, Hisako
: APPLICANT: Tamai, Yukio
: APPLICANT: Iwamatsu, Akihiro
: APPLICANT: Kato, No. 6001590uo
: APPLICANT: Sakai, Yasuyoshi
: TITLE OF INVENTION: PROMOTER/TERMINATOR FOR CANDIDA BOIDINI
: TITLE OF INVENTION: FORMATE DEHYDROGENASE GENE
: NUMBER OF SEQUENCES: 51
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Foley & Lardner
: STREET: 3000 K Street, N.W., Suite 500
: CITY: Washington
: STATE: D.C.
: COUNTRY: USA
: ZIP: 20007-5109
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patentin Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/817,926
: FILING DATE: 09-MAY-1997
: CLASSIFICATION: 435
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: PCT/JP96/02597
: FILING DATE: 12-SEP-1996
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: JP 234133/1995
: FILING DATE: 12-SEP-1995
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: JP 42536/1996
: FILING DATE: 29-FEB-1996
: ATTORNEY/AGENT INFORMATION:
: NAME: Bent, Stephen A.
: REGISTRATION NUMBER: 29, 768
: REFERENCE/DOCKET NUMBER: 081356/0112
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (202)672-5300
: TELEFAX: (202)672-5399
: TELETYPE: 904136
: INFORMATION FOR SEQ ID NO: 27:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 4818 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: double
: TOPOLOGY: linear
: MOLECULE TYPE: DNA (genomic)
: ORIGINAL SOURCE:
: ORGANISM: Candida boidinii
: US-08-817-926-27

Query Match      2.0%; Score 62.6; DB 3; Length 4818;
Best Local Similarity 44.4%; Pred. No. 0.00066;
Matches 338; Conservative 0; Mismatches 419; Indels 4; Gaps 2;

```

```

QY 1705 cgaagtaaaaaaatttcttatttctgcgagtcgcatgaagataaccatgactt 1764
      || || || || || || || || || || || || || || || || || || ||
DB 4199 ATTATTATTATTATTATTCTGCTGCAACGTTTAAAAAGAGTAGTATTAACTTAACACTGTT 4258
QY 1765 aattttgaaaaatgtaacccttact--catagattaatcacctgatttctgtg 1821
      || || || || || || || || || || || || || || || || || || ||
DB 4259 CTCAGTTAATAGTGAATTAACACTGATTTAAATACCTTTTTCCTGCAATTTTTCAA 4318
QY 1822 ccatatgacgccttacaactcgtatgtaacatcttttcttcgcaaatatlaagaa 1881
      || || || || || || || || || || || || || || || || || || ||
DB 4319 TTATTCATTTTGGTAATTTGGTAAATTAATGCTGAATTAATAAAAAAAATTTCTGTTT 4378
QY 1882 atcaatgcctactcaatagaagaacagcgtagatattacatttaattlaagacaa 1941
      || || || || || || || || || || || || || || || || || || ||
DB 4379 CTTTGTGTTTATTTCTGTTTAAATTTTGTGATCAATCTTTAAATTTAGTTTAA 4438
QY 1942 attttgaaaaatgtaataatttcaacaatattatlaaataatgtagcctaatglat 2001
      || || || || || || || || || || || || || || || || || || ||
DB 4439 ATTTTATATTATTTCTAATTTAATTAATTAATGTTGTAATTAATTAATTAATTAAT 4498
QY 2002 ttccatgcttcttaaaatatlttttlatattagttataaatacatgaaccaata 2061
      || || || || || || || || || || || || || || || || || || ||
DB 4499 TTCTAATTTTGTGAACAAACGTTTAAATTTAGATATTAAATTTAATTTAATTA 4558
QY 2062 atagtggtgaatccaatatcccatlaatttttgaacctacaatatlaatat 2121
      || || || || || || || || || || || || || || || || || || ||
DB 4559 TTGTGTTTAAATTCGTATACCATTTTCTTTTACTGTAATTTAATTTAATTAATTA 4618
QY 2122 ttagcaataacaatgcatagaaagttccaaaaaaatttggtaacagaacatcccaa 2181
      || || || || || || || || || || || || || || || || || || ||
DB 4619 TTATTTATCGAAAAAACCTTAAAAAGACTGCTGAATTTTATTTGAATTAACGAAAAA 4678
QY 2182 tttttttttttaaagacaagaataacagaatagaacacttttctgtcgtgaatgaa 2241
      || || || || || || || || || || || || || || || || || || ||
DB 4679 TGACATATATTAACAAAAAATAATTAATAAGAAATTAATAAACAAGTGTGATTTTACTTT 4738
QY 2242 gtaagtaataacataagcaaatlttaaaaaaatataataaag 2282
      || || || || || || || || || || || || || || || || || || ||
DB 4739 TAACTAAGTCAAGTCTAATAATTTCTAATTTTCTTTTAAAG 4779

RESULT 9
US-07-867-106-2/c
: Sequence 2, Application US/07867106
: Patent No. 5389526
: GENERAL INFORMATION:
: APPLICANT: Slade, Martin B
: APPLICANT: Slade, Andy C M
: APPLICANT: Williams, Keith L
: TITLE OF INVENTION: Improved Plasmid Vectors for Cellular
: NUMBER OF SEQUENCES: 19
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Woodcock Washburn Kurtz Mackiewicz & No. 5389526ris
: STREET: One Liberty Place 46th Floor
: CITY: Philadelphia
: STATE: PA
: COUNTRY: USA
: ZIP: 19103
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: OPERATING SYSTEM: IBM PC compatible
: SOFTWARE: Patentin Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/07/867,106
: FILING DATE: 19920625
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: AU PJ 7187
: APPLICATION NUMBER: PCT/AU90/00530
: FILING DATE: 02-NOV-1989
: ATTORNEY/AGENT INFORMATION:

```


RESULT 11
US-08-998-416-288/c
: Sequence 288, Application US/08998416
: Patent No. 6239264

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
84

CORRESPONDENCE ADDRESS:
ADDRESSEE: No. 6239264rtis Corporation
STREET: 3054 Cornwallis Road
CITY: Research Triangle Park
STATE: No. 6239264rt Carolina
COUNTRY: USA
ZIP: 27709
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.3
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/998,416
FILING DATE: 24-DEC-1997
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: CH 0016/97
FILING DATE: 31-DEC-1996
ATTORNEY/AGENT INFORMATION:
NAME: Meigs, J. Timothy
REGISTRATION NUMBER: 38,241
REFERENCE/DOCKET NUMBER: Pp/5-30306/A/CCG1976
TELECOMMUNICATION INFORMATION:
TELEPHONE: 919-541-8587
TELEFAX: 919-541-8689
INFORMATION FOR SEQ ID NO: 288:
SEQUENCE CHARACTERISTICS:
LENGTH: 837 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
ORIGINAL SOURCE:
ORGANISM: PAG1241RP
US-08-998-416-288

| | | | | |
|-----------------------|--------|-------------------|-------|-------------|
| Query Match | 1.9%; | Score 60.6; | DB 4; | Length 837; |
| Best Local Similarity | 49.1%; | Pred. No. 0.0011; | | |

RESULT 12
US-08-446-855A-1/c

? Sequence 1, Application US/08446855A
 ? Patent No. 5849573
 ?
 ? GENERAL INFORMATION:
 ?
 ? APPLICANT: Stewart, Thomas S
 ? APPLICANT: Flores, Maria V
 ? APPLICANT: O'Sullivan, William J
 ? TITLE OF INVENTION: Nucleotide sequence encoding carbamoyl
 ? TITLE OF INVENTION: phosphate synthetase II
 ? NUMBER OF SEQUENCES: 2
 ? CORRESPONDENCE ADDRESS:
 ? ADDRESSER: Nixon & Vanderhye PC
 ? STREET: 1100 No. 5849573th Glebe Road, 8th Floor
 ? CITY: Arlington
 ? STATE: Virginia
 ? COUNTRY: USA
 ? ZIP: 22201-4714
 ?
 ? COMPUTER READABLE FORM:
 ?
 ? MEDIUM TYPE: Floppy disk
 ? COMPUTER: IBM PC compatible
 ? OPERATING SYSTEM: PC-DOS/MS-DOS
 ? SOFTWARE: PatentIn Release #1.24
 ? CURRENT APPLICATION DATA:
 ? APPLICATION NUMBER: US/08/446,855A
 ? FILING DATE: 06-Jul-1995
 ?
 ? CLASSIFICATION: 435
 ?
 ? ATTORNEY/AGENT INFORMATION:
 ? NAME: Mitchard, Leonard C
 ? REGISTRATION NUMBER: 29,009
 ? REFERENCE/DOCKET NUMBER: 47-80
 ? TELECOMMUNICATION INFORMATION:
 ? TELEPHONE: 703-816-4000
 ? TELEFAX: 703-816-4100
 ?
 ? INFORMATION FOR SEQ ID NO: 1:
 ? SEQUENCE CHARACTERISTICS:
 ? LENGTH: 8920 base pairs
 ? TYPE: nucleic acid
 ? STRANDEDNESS: single
 ? TOPOLOGY: linear
 ?
 ? MOLECULE TYPE: genomic
 ?
 ? US-08-446-855A-1

| | | | |
|----|------|---|------|
| Oy | 1976 | atlaaaatagaagcctaatagtaattcciaagtcttcaaaaataftttttatat | 2035 |
| Dd | 142 | TTTTTATATTAAATATTTTTAATTAAATTTTAAATTTAAATAAATA | 201 |
| Oy | 2036 | agtlataaatacatlgaaccaataagtcgttgynaaticcaatctcatnaat | 2095 |
| Dd | 202 | TTTTTATATTAAATATTTTTAATTAAATTTTAAATTTAAATAAATA | 261 |
| Oy | 2096 | tlttgaaatcaccaatttaat--atttaglcaataacaatgcgatgaaagtcca | 2153 |
| Dd | 262 | TTTTTATATTAAATGTTTTAAATTAATAATTTAAATTTAAATGTTTAAATTACATA | 321 |
| Oy | 2154 | aaaaattcgtaaagaaactccaattctttcttaacygaacaagaataacaga | 2213 |
| Dd | 352 | TTTTTATATTAAATGTTTTTAATTACATATTTTAATTTAAATGTTTAAATTACATA | 381 |
| Oy | 2214 | tagaaaactatctgtctgygaatggaagtagbaatalacatalaacgaactttaaaaa | 2273 |
| Dd | 382 | TTTTTATATTAAATGTTTTTAATTACATATTTTAATTTAAATGTTTAAATTACATA | 441 |
| Oy | 2274 | ttataagaacctaaagcgcgtcaaagtatgtaactgaagglytaaltaataagcgt | 2333 |
| Dd | 442 | TTTTTATATTACATATTTTAT-AAAGATTTTAAATTACATATTTTAATTAATTAAGTAT | 500 |
| Oy | 2334 | glgcgactcagaatctgggaacaacaatgaaacggaatlaaaatlataacttlaaataaa | 2393 |
| Dd | 501 | TATATATTACATATTTTAAATTAAAGATTTTAAATTACATATTTTAAATTAATTCAATATT | 560 |
| Oy | 2394 | taaaaat 2400 | |
| Dd | 561 | TATTAAT 567 | |

RESULT 15
 ; Sequence 1, Application US/08947823
 ; Patent No. 6114605
 ; GENERAL INFORMATION:
 APPLICANT: Williamson, Valerie M.
 APPLICANT: Kaloshian, Isgouhi
 APPLICANT: Yaghooobi, Jafar
 APPLICANT: Bodeau, John
 APPLICANT: Milligan, Stephen
 TITLE OF INVENTION: Procedures and Materials for Confering
 NUMBER OF SEQUENCES: 5
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Townsend and Townsend and Crew LLP
 STREET: Two Embarcadero Center, Eighth Floor
 CITY: San Francisco
 STATE: California
 COUNTRY: USA
 ZIP: 94111-3834
 ; COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: PatentIn Release #1.0, Version #1.30
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/947,823
 FILING DATE: 09-OCT-1997
 CLASSIFICATION: 800
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: PCT/US97/18802
 FILING DATE: 09-OCT-1997
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 60/028,191
 FILING DATE: 10-OCT-1996
 ATTORNEY/AGENT INFORMATION:
 NAME: Bastian, Kevin L.
 REGISTRATION NUMBER: 34,774
 REFERENCE/DOCKET NUMBER: 023070-07021005
 TELECOMMUNICATION INFORMATION:

```

; TELEPHONE: (415) 576-0200
; TELEFAX: (415) 576-0300
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 51952 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
US-08-947-823-1

```

| Query Match | Similarity | 1.8% | Score | 59.2 | DB | 3 | Length | 51952 |
|-------------|------------|---|-------|------------|--------|--------|--------|-------|
| Best Local | Similarity | 43.7% | Pred. | No. | 0.0052 | | | |
| Matches | 410 | Conservative | 0 | Mismatches | 518 | Indels | 10 | Gaps |
| QY | 1524 | aactataatgctcttcttaatgcatctttaacatattatgttgaattttaa | 1583 | | | | | |
| Db | 31954 | AACAATACATCAAAATCTTTTCTTAATGTCATGTATATCTCATGATGCACATTTA | 32013 | | | | | |
| QY | 1584 | taagcgaacttglytaacatlaacaatattatattagatactagatgattatccaa | 1643 | | | | | |
| Db | 32014 | ATTTAGACAGAAATTTCTGTGATTTTTTTTTGTAGTGTGATCTTCATTTAACCAA | 32072 | | | | | |
| QY | 1644 | atacaacttggatgtttaaactaatcttcttcttctcctaagratataatca | 1703 | | | | | |
| Db | 32074 | AAATACATCTTAACACATTTTAGTAATTTTAATATATTTTAAGGGCGAACAAATTAATATAT | 32133 | | | | | |
| QY | 1704 | tcgaggtlaaaaaagtttgccttattcttcgcgatcagaaagataaactaatgact | 1763 | | | | | |
| Db | 32134 | AATTATGATGAAATTTTGAGAGAAATTCATACCTCTGTACTTTTAATTTGCACATTA | 32199 | | | | | |
| QY | 1764 | taatttttggaaaatglaaccctttactcatagatlaaattacglatgtttgttgc | 1823 | | | | | |
| Db | 32194 | AAAGATTTTTTTTATATATGTTTTTAAAGAGAGTTTCATTAATAGCTTTTTCAAA | 32253 | | | | | |
| QY | 1824 | ataatgacgcctctcaactcgaatgtagcaatctttctgcgaatataataggat | 1883 | | | | | |
| Db | 32254 | CATTAACAAGTCTTAAGCTTGCTTGACATGTTTTTGAAGAAATTTTGA--ACCTTAT | 32311 | | | | | |
| QY | 1884 | tcaatgtaactcaatagaagaacagctgagtattacatttaattaaagacaaat | 1943 | | | | | |
| Db | 32312 | TTRAATTAATAATCTGTTTGATGTATATTTTATCTCPAAAATCTTTGAAAATTTAT | 32377 | | | | | |
| QY | 1944 | tttggaaaaatgtaaatcttcaacaatattaaatlatgagcctaataatgatt | 2003 | | | | | |
| Db | 32372 | TGCGTAAACATGTTTTTAATATTTGTTACGATTAAGACAAATTCGATCTTGATTAT | 32433 | | | | | |
| QY | 2004 | cctatgttcttaaatatattttttttatatatttgattataatca---ttaagaacaa | 2060 | | | | | |
| Db | 32432 | TAAAAATCATCTTTAAATATGATATATCAAAATTTTTTAAATTAATTAATCTTTTAAAAACAT | 32491 | | | | | |
| QY | 2061 | aatagtttggatgtcaaatatctcatlaaatttttttgaatctcaaatattcaat | 2120 | | | | | |
| Db | 32492 | AATTAAGCTAACTTTGTTTGTGACATGTTTTTGAAGGAATTTTAAACAATTTTAAATA | 32555 | | | | | |
| QY | 2121 | tttgaacaatacaatgcatagaagtctcoaaaaaaatlttlaacagaaaacttccaa | 2180 | | | | | |
| Db | 32552 | AATGTGTTGAATAGTTAATATTTGTTCTCTTAATAAACACTTTAAGAAATCTTATGTAA | 32611 | | | | | |
| QY | 2181 | atttttttttlaatggacaagaataacacagatagaaaaactatttgttggaaatgga | 2240 | | | | | |
| Db | 32612 | CATATCTTATATTTCTTACAGAAATTTGACCACTTAATTCGTATCTTCATTTGAATA | 32677 | | | | | |
| QY | 2241 | agtagaataataaataagcaaaattttaaanaattatataagcctta-----cgagctc | 2295 | | | | | |
| Db | 32672 | ATCTTTAAATGACGTATCAACATTTTTTAATTAATTTGAAGAAATTTAAATTAATACATGAC | 32733 | | | | | |
| QY | 2296 | aagatgtatcatcagtaagtgtaaatlaaataagcatgtygcgatccgaattggagaaa | 2355 | | | | | |
| Db | 32732 | AAAGAAAAAACAATGTTTGGACATTAATTAAGAAATTAATGATTCGATTAATGATTCATCAAA | 32791 | | | | | |
| QY | 2356 | caatgaaaacggaatataaatatlaaactttaaataataaaaaatttggtaaagtgt | 2415 | | | | | |

```

Db 32792 AGATTAAAAATAAAAATCAATTTTAAATAATAATATTTCTTGATATATAATATG 32851
Qy 2416 ttctgactatlgagggaacaaagacatgcacaa 2453
Db 32852 ATAACTATTTAACAGGGGAAATTTGACNAGAAAAA 32889

```

Search completed: October 17, 2001, 19:13:08
 Job time: 15663 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 17, 2001, 14:40:40 ; Search time 5433.22 Seconds
(without alignments)
5570.921 Million cell updates/sec

Title: US-09-502-426-1_COPY_1_3202
Perfect score: 3202

Sequence: 1 atgtgggtattatattgttg;.....gagagagagaactagctcc 3202

Scoring table: IDENTITY NUC
Gapop 10.0, Gapext 1.0

Searched: 10228115 segs, 4726426750 residues

Total number of hits satisfying chosen parameters: 20456230

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:*

- 1: gb_est1:*
- 2: gb_est2:*
- 3: gb_est3:*
- 4: gb_est4:*
- 5: gb_est5:*
- 6: gb_est6:*
- 7: gb_est7:*
- 8: gb_est8:*
- 9: gb_est9:*
- 10: gb_est10:*
- 11: gb_est11:*
- 12: gb_est12:*
- 13: gb_est13:*
- 14: gb_est14:*
- 15: gb_est15:*
- 16: gb_est16:*
- 17: gb_est17:*
- 18: gb_est18:*
- 19: gb_est19:*
- 20: gb_est20:*
- 21: gb_est21:*
- 22: gb_est22:*
- 23: gb_est23:*
- 24: gb_est24:*
- 25: gb_est25:*
- 26: gb_est26:*
- 27: gb_est27:*
- 28: gb_est28:*
- 29: gb_est29:*
- 30: gb_est30:*
- 31: gb_est31:*
- 32: gb_est32:*
- 33: gb_est33:*
- 34: gb_est34:*
- 35: gb_est35:*
- 36: gb_est36:*
- 37: gb_est37:*
- 38: gb_est38:*
- 39: gb_est39:*
- 40: gb_est40:*
- 41: gb_est41:*
- 42: gb_est42:*
- 43: gb_est43:*

- 44: em_esthum10:*
- 45: em_esthum11:*
- 46: em_esthum12:*
- 47: em_esthum13:*
- 48: em_esthum14:*
- 49: em_esthum15:*
- 50: em_esthum16:*
- 51: em_esthum17:*
- 52: em_esthum18:*
- 53: em_esthum19:*
- 54: em_esthum20:*
- 55: em_esthum21:*
- 56: em_esthum22:*
- 57: em_esthum23:*
- 58: em_esthum24:*
- 59: em_esthum25:*
- 60: em_esthum26:*
- 61: em_esthum27:*
- 62: em_esthum28:*
- 63: em_esthum29:*
- 64: em_esthum30:*
- 65: em_esthum31:*
- 66: em_esthum32:*
- 67: em_esthum33:*
- 68: em_esthum34:*
- 69: em_esthum35:*
- 70: em_esthum36:*
- 71: em_esthum37:*
- 72: em_esthum38:*
- 73: em_esthum39:*
- 74: em_esthum40:*
- 75: em_esthum41:*
- 76: em_esthum42:*
- 77: em_esthum43:*
- 78: em_esthum44:*
- 79: em_esthum45:*
- 80: em_esthum46:*
- 81: em_esthum47:*
- 82: em_esthum48:*
- 83: em_esthum49:*
- 84: em_esthum50:*
- 85: em_esthum51:*
- 86: em_esthum52:*
- 87: em_esthum53:*
- 88: em_esthum54:*
- 89: em_esthum55:*
- 90: em_esthum56:*
- 91: em_esthum57:*
- 92: em_esthum58:*
- 93: em_esthum59:*
- 94: em_esthum60:*
- 95: em_esthum61:*
- 96: em_esthum62:*
- 97: em_esthum63:*
- 98: em_esthum64:*
- 99: em_esthum65:*
- 100: em_esthum66:*
- 101: em_esthum67:*
- 102: em_esthum68:*
- 103: em_esthum69:*
- 104: em_esthum70:*
- 105: em_esthum71:*
- 106: em_esthum72:*
- 107: em_esthum73:*
- 108: em_esthum74:*
- 109: em_esthum75:*
- 110: em_esthum76:*
- 111: em_esthum77:*
- 112: em_esthum78:*
- 113: em_esthum79:*
- 114: em_esthum80:*
- 115: em_esthum81:*
- 116: em_esthum82:*

```
117: gb_est148:*
118: gb_est149:*
119: gb_est150:*
120: gb_est151:*
121: gb_est152:*
122: gb_est153:*
123: gb_est154:*
124: gb_est155:*
125: gb_est156:*
126: gb_est157:*
127: gb_est158:*
128: gb_est159:*
129: gb_est160:*
130: gb_est161:*
131: gb_est162:*
132: gb_est163:*
133: gb_est164:*
134: gb_est165:*
135: gb_est166:*
136: gb_est167:*
137: gb_est168:*
138: gb_est169:*
139: gb_est170:*
140: gb_est171:*
141: gb_est172:*
142: gb_est173:*
143: gb_est174:*
144: gb_est175:*
145: gb_est176:*
146: gb_est177:*
147: gb_est178:*
148: gb_est179:*
149: gb_est180:*
150: gb_est181:*
151: gb_est182:*
152: gb_est183:*
153: gb_est184:*
154: gb_est185:*
155: gb_est186:*
156: gb_est187:*
157: gb_est188:*
158: gb_est189:*
159: gb_est190:*
160: gb_est191:*
161: gb_est192:*
162: gb_est193:*
163: gb_est194:*
164: gb_est195:*
165: gb_est196:*
166: gb_est197:*
167: gb_est198:*
168: gb_est199:*
169: gb_est200:*
170: gb_est201:*
171: gb_est202:*
172: gb_est203:*
173: gb_est204:*
174: gb_est205:*
175: gb_est206:*
176: gb_est207:*
177: gb_est208:*
178: gb_est209:*
179: gb_est210:*
180: gb_est211:*
181: gb_est212:*
182: gb_est213:*
183: gb_est214:*
184: gb_est215:*
185: gb_est216:*
186: gb_est217:*
187: gb_est218:*
188: gb_est219:*
189: gb_est220:*

190: gb_est110:*
191: gb_est111:*
192: gb_hic:*
193: em_gss_fun:*
194: em_gss_hum1:*
195: em_gss_hum2:*
196: em_gss_hum3:*
197: em_gss_hum4:*
198: em_gss_hum5:*
199: em_gss_hum6:*
200: em_gss_hum7:*
201: em_gss_hum8:*
202: em_gss_hum9:*
203: em_gss_inv1:*
204: em_gss_inv2:*
205: em_gss_inv3:*
206: em_gss_inv4:*
207: em_gss_inv5:*
208: em_gss_inv6:*
209: em_gss_inv7:*
210: em_gss_inv8:*
211: em_gss_inv9:*
212: em_gss_inv10:*
213: em_gss_inv11:*
214: em_gss_inv12:*
215: em_gss_inv13:*
216: em_gss_inv14:*
217: em_gss_inv15:*
218: em_gss_inv16:*
219: em_gss_inv17:*
220: em_gss_inv18:*
221: em_gss_inv19:*
222: em_gss_inv20:*
223: em_gss_inv21:*
224: em_gss_inv22:*
225: em_gss_inv23:*
226: em_gss_inv24:*
227: em_gss_inv25:*
228: em_gss_inv26:*
229: em_gss_inv27:*
230: em_gss_inv28:*
231: em_gss_inv29:*
232: em_gss_inv30:*
233: em_gss_inv31:*
234: em_gss_inv32:*
235: em_gss_inv33:*
236: em_gss_inv34:*
237: em_gss_inv35:*
238: em_gss_inv36:*
239: em_gss_inv37:*
240: em_gss_inv38:*
241: em_gss_inv39:*
242: em_gss_inv40:*
243: em_gss_inv41:*
244: em_gss_inv42:*
245: em_gss_inv43:*
246: em_gss_inv44:*
247: em_gss_inv45:*
248: em_gss_inv46:*
249: em_gss_inv47:*
250: em_gss_inv48:*
251: em_gss_inv49:*
252: em_gss_inv50:*
253: em_gss_inv51:*
254: em_gss_inv52:*
255: em_gss_inv53:*
256: em_gss_inv54:*
257: em_gss_inv55:*
258: em_gss_inv56:*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

[illegible]

ALIGNMENTS

| | |
|------------|---|
| RESULT | 1 |
| CNS00EVL/c | |
| LOCUS | CNS00EVL 1101 bp DNA GSS 04-JUN-1999 |
| DEFINITION | Drosophila melanogaster genome survey sequence T7 end of BAC: |
| | BACR2B823 of RPi-98 library from Drosophila melanogaster (fruit |
| | fly), genomic survey sequence. |
| ACCESSION | AL069706 |
| VERSION | AL069706.1 GI:4949849 |
| KEYWORDS | GSS. |
| SOURCE | Fruit fly. |
| ORGANISM | Drosophila melanogaster |
| | Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta; |
| | Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; |
| | Muscomorpha; Ephydroidea; Drosophilidae; Drosophila. |
| REFERENCE | 1 (bases 1 to 1101) |
| AUTHORS | Genoscope. |
| TITLE | Direct Submission |

JOURNAL

Submitted (02-JUN-1999) Genoscope - Centre National de Séquençage
BP 191 91006 Evry cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
Web : www.genoscope.cns.fr)
Determination of this BAC-end sequence was carried out as part of a
collaboration with the Berkeley Drosophila Genome Project (BDGP).
The BDGP is constructing a physical map of the Drosophila
melanogaster genome using these BACs. For further information
please see <http://www.fruitfly.org> The BDGP Drosophila
melanogaster BAC library was prepared by Kazuhiro Oseegawa and
Aaron Memminger in Pierre de Jong's laboratory in the Department of
Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
NY. The library is named RPCL-98 and was constructed by partial
EcoRI digestion of Drosophila DNA provided by the BDGP from the
isogenic strain Y2: cm bw sp. the same strain used for the BDGP's
p1 and EST libraries. A more detailed description of the library
and how to order individual BAC clones, the entire library, or
filters for hybridization from the BACPAC Resource Center can be
found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.

FEATURES

```

Source
1. .1101
/organism="Drosophila melanogaster"
/db_xref="taxon:7227"
/clone_1b="RPCI_98"
/clone="BACR29B23"
/note="end : 17"
419 a 91 c 60 g 259 t 232 others
BASE COUNT
ORIGIN

```

ORIGIN

| | | | | |
|-----------------------|-------------------|--------------------|-----------|--------------|
| Query Match | 3.5%; | Score 113.2; | DB 219; | Length 1101; |
| Best Local Similarity | 35.8%; | Pred. No. 3.3e-09; | | |
| Matches 232; | Conservative 118; | Mismatches 297; | Indels 1; | Gaps 1 |

| | | | |
|----|------|--|-------|
| Oy | 1625 | lagaatagatcatcccaatacaataacttggctgcttgaacttaacatccctgcttctcc | 1864 |
| Db | 1100 | TTMTATWTMTTWTMTTMTATMTATMTATMTATMTATMTATMTATMTAT | 10411 |
| Oy | 1685 | acggtatacaatatataccatccgcggtaaaaaagtttctctatttccgcgatgcga | 17444 |
| Db | 1040 | WATATATTATMTTWTATAAATATMTMTTATATAATATATATMTATMTATMTATAA | 981 |
| Oy | 1745 | aggaataacccaagacttaacttcttggaaaaatgaacccttaccatagataat | 1804 |
| Db | 960 | WTATAMATMTMAATMAATMTMATMTTATATAMAAATATMAAMMTWTATATTTT | 921 |
| Oy | 1805 | accgatgcttcttgccataatgacagccctacaactgtagtcaagtcatacttctc | 1864 |
| Db | 920 | TTTWTTTTMTATMTTATATMTMTMTMAAAMAMAMMTATTTTWTATMTATMTATA | 861 |
| Oy | 1865 | caaatatnaattaggaatccaatgcctactcaatcagaagaacagctgaglatcat | 19244 |
| Db | 860 | TTATMTTMTTWTMTATMTMTATATATATMTATMTTWTMTTATMTATTTATMTWT | 801 |
| Oy | 1925 | ttaactttaagaacaaatcttggaaaaatgctataacttccaacatatcttaata | 19844 |
| Db | 800 | WTTWTATTTATATATATMTMTMTMTMTMTMTMAAMTATATATMTATATATMTATMTT | 741 |
| Oy | 1985 | tgaagccataaagtatactctat-gtcttaaaatcttcttatacttatactagataa | 20434 |
| Db | 740 | WTTTATTTTWTTTTAAMTTATMTATMTATMTTWTMTATWTTTTTTTTTTTTTTTT | 681 |
| Oy | 2044 | atacatatagaaccaataaagttgctggaatcaaatcatctccatnaatcttggaa | 21034 |
| Db | 680 | TTATTTTATATMTMAATMTATTTTAAATATMTATTTATATTTATAAATTTATMTAA | 621 |
| Oy | 2104 | tctcaaatatataatcttagtccaataacaatgatgaaagttcccaaaaaacttctg | 21634 |
| Db | 620 | AAATATTTTATTTATTTAATMTAMATTTATMTATTAATATAMTWTTAATTTTATATTA | 561 |
| Oy | 2164 | ttaacagaacttccaaatcttcttcttatactgaaacagaataacagatagaaacta | 22234 |
| Db | 560 | TTMTTATTAATATATWTTTTTTTTTTATMTATMTMTMAAAMATATMAAMMTMAAATMTMT | 501 |

| RESULT | 2 |
|--|---|
| CNS00EVL | 1101 bp DNA GSS 04-JUN-1999 |
| LOCUS | Drosophila melanogaster genome survey sequence T7 end of BAC: BACR2B23 of RPCT-98 library from Drosophila melanogaster (fruit fly), genomic survey sequence. |
| DEFINITION | AL069706 GI:4949849 |
| ACCESSION | AL069706.1 |
| VERSION | GSS. |
| KEYWORDS | fruit fly. |
| SOURCE | Drosophila melanogaster |
| ORGANISM | Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila. |
| REFERENCE | 1 (bases 1 to 1101) |
| AUTHORS | Genoscope. |
| TITLE | Direct Submission |
| JOURNAL | Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr - web : www.genoscope.cns.fr) |
| COMMENT | Determination of this BAC-end sequence was carried out as part of a collaboration with the Berkeley Drosophila Genome Project (BDGP). The BDGP is constructing a physical map of the Drosophila melanogaster genome using these BACs. For further information please see http://www.fruitfly.org/TheBDGP/drosophila_melanogaster_BAC_library was prepared by Kazutoyo Osoegawa and Aaron Mammoss in Pieter de Jong's laboratory in the Department of Cancer Genetics at the Roswell Park Cancer Institute in Buffalo, NY. The library is named RPCT-98 and was constructed by partial EcoRI digestion of Drosophila DNA provided by the BDGP from the isogenic strain Y2: cn bw sp, the same strain used for the BDGP's p1 and EST libraries. A more detailed description of the library and how to order individual BAC clones, the entire library, or filters for hybridization from the BACPAC Resource Center can be found at http://bacpac.med.buffalo.edu/drosophila_bac.htm . |
| FEATURES | Location/Qualifiers |
| source | 1..1101 |
| | /organism="Drosophila melanogaster" |
| | /db_xref="taxon:7227" |
| | /clone_1db="RPCT-98" |
| | /clone="BACR2B23" |
| | /note="end : T7" |
| BASE COUNT | 419 a 91 c 60 g 299 t 232 others |
| ORIGIN | |
| Query Match | 3.5%: Score 110.8; DB 219; Length 1101; |
| Best Local Similarity | 34.5%: Pred. No. 8e-09; |
| Matches 223: Conservative 128; Mismatches 290; Indels 5; Gaps 1; | |
| QY 1513 | taacacatcgcaacattaaatgtcttccttaagcactcttaacatattatgttag 1572 |
| Db 461 | TTWTMMMMMMMAATTTWMAAAMAATTAATTAATAAAMAAMMMMAATTTTMMWMTMA 520 |
| QY 1573 | ctggaacttaagaagcgcaactgttaacattacaatatattatagatagtagt 1632 |
| Db 521 | TTTWTMMWMTWTATTAATAAAMAAMAATAATTTAAAMMAATATTAAMAAMW 580 |
| QY 1633 | gattatccaataactcttggttggttaactaactctgttctcccaaggtata 1692 |
| Db 581 | TATATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 640 |
| QY 1693 | aattatacatcgaggaataaagaatttgcattatctcgcatgcagagataaa 1752 |
| Db 641 | TTTAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 700 |
| QY 1753 | cttaaatgacttaattcttgaataatgtaacccttctacatagataattaccgta 1812 |

| | | | | |
|-------------------|---|---|--|------|
| Df | 701 | | AAAAAATWAAATAWTMTAMTATAMTTAAAANAATTAAAAAMAATAWTMAWAATAATMTMA | 760 |
| Oy | 1813 | tltttgttgccatcatagacgccttcatacactgtagtacaatttttcctccaatat | 1872 | |
| Dd | 761 | TATATATMTTWTAATAAWATWWAAWWWTTATAVA---- | TAAATATAWAAMAATAATAATMWATA | 815 |
| Oy | 1873 | aaatagggaattcaatcgacctcataagaagaacagctgagtataaccattaact | 1932 | |
| Dd | 816 | MATMNAANAAWAATATMMATMTATATMMATATAWANAAAAAAAAATTAATATTAATMTATMANA | 875 | |
| Oy | 1933 | aaagacaaaatttgtaaaaatgltalaattcctaacaatlataltaaacatgatgcot | 1992 | |
| Dd | 876 | AAAAATWAAWTMTWTTTTWWNNNNTATFAAANAMATMANAAAAAANAANAANA | 935 | |
| Oy | 1993 | ataagctttcccatagtctccttaaabaatttttttatattagttlataaactact | 2052 | |
| Dd | 936 | MMMMTATMTTTTATTAATAAATFTATMMATTTTMMATTMWTATMTATATATATATW | 995 | |
| Oy | 2053 | gaaccaataatagttggigaattccaataatgccccaataatatttttgaaattcaa | 2112 | |
| Dd | 996 | WTAAWTAAATATATTATTAATAAATATATTATTTTAAAAAAWTATATATATNM | AAMWTAMANATATAM | 1055 |
| Oy | 2113 | tattatattatgcatcacatgcatacgataagtcataaaaaaaaa | 2158 | |
| Dd | 1056 | AMWAATTATMTATATATATTAATAWATAWMAAANAATAWATAMAA | 1101 | |
| RESULT_3 LOCUS | CNS0021J/c | 1101 bp DNA GSS | 03-JUN-1999 | |
| DEFINITION | Drosophila melanogaster genome survey sequence TET3 end of BAC # | BACR05N11 of Rpci-98 library from Drosophila melanogaster (fruit fly), genomic survey sequence. | | |
| ACCESSION | AL061936 | | | |
| VERSION | ALI061936.1 | GI:4940214 | GSS. | |
| SOURCE | fruit fly, | | | |
| ORGANISM | Drosophila melanogaster | | | |
| | Euarystota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta; | | | |
| | Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; | | | |
| | Muscomorpha; Ephydroidea; Drosophilidae; Drosophila. | | | |
| REFERENCE | 1 (bases 1 to 1101) | | | |
| AUTHORS | Genoscope. | | | |
| TITLE | Direct Submission | | | |
| JOURNAL | Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage BP 191 91006 Evry cedex - FRANCE (E-mail : seqref@genoscope.cns.fr - web : www.genoscope.cns.fr) | | | |
| COMMENT | determination of this BAC-end sequence was carried out as part of a collaboration with the Berkeley Drosophila Genome Project (BDGP). The BDGP is constructing a physical map of the Drosophila melanogaster genome using three BACS. For further information please see http://www.fruitfly.org The BDGP Drosophila melanogaster BAC library was prepared by Kazutoyo Osoegawa and Aaron Mammoser in Pieter de Jong's laboratory in the Department of Cancer Genetics at the Roswell Park Cancer Institute in Buffalo, NY. The library is named Rpci-98 and was constructed by partial EcoRI digestion of Drosophila DNA provided by the BDGP from the isogenic strain y2; cn bw sp, the same strain used for the BDGP's pI and EST libraries. A more detailed description of the library and how to order individual BAC clones, the entire library, or filters for hybridization from the BACPAC Resource Center can be found at http://bacpac.med.buffalo.edu/drosophila_bac.htm. | | | |
| FEATURES | Location/Qualifiers | | | |
| source | 1..1101 /organism="Drosophila melanogaster" /db_xref="taxon:7227"/clone_lib="RPCI-98"/clone="BACR05N11"/note="end : TET3" | | | |
| BASE COUNT | 631 a 7 c 28 g 289 t | | 146 others | |
| ORIGIN | | | | |

TITLE Direct Submission
KEYWORDS Genoscope - Centre National de Séquençage
ABSTRACT Submitted (02-JUN-1999)
ADDRESS BP 191 91006 EVRY cedex - FRANCE (E-mail : segreff@genoscope.cns.fr)
URL - web : www.genoscope.cns.fr

Oy 1762 tttaatttttgaataatgtaaccctttaactcataagattaattaacgatgtlitttgty 1823
||| |||||| | | |||| : | | |||| |
Db 1046 ttttttttttaaataataatatatttatwtttatatattttattttttttttttta 987

Query Match 2.98; Score 94; DB 219; Length 1101;

| | |
|---------------------------|---|
| ACCESSION | fly), genomic survey sequence. |
| VERSION | AJ064091 |
| KEYWORDS | AI064091.1 GI:4941847 |
| SOURCE | GSS. |
| ORGANISM | Fruit fly. |
| REFERENCE | Drosophila melanogaster |
| AUTHORS | Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta; |
| TITLE | Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; |
| JOURNAL | Muscomorpha; Ephydroidea; Drosophilidae; Drosophila. |
| | 1 (bases 1 to 1101) |
| | Genoscope. |
| | Direct Submission |
| | Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage : |
| | BP 191 91006 Evry cedex - FRANCE (E-mail : seqrefgenoscope.cns.fr |
| | - Web : www.genoscope.cns.fr) |
| | Determination of this BAC-end sequence was carried out as part of a |
| | collaboration with the Berkeley Drosophila Genome project (BDGP). |
| | The BDGP is constructing a physical map of the Drosophila |
| | melanogaster genome using these BACs. For further information |
| | please see http://www.fruitfly.org/The_BDGP_Drosophila |
| | melanogaster_BAC_library.html |
| | Aeron Mammoser in Pieter de Jong's laboratory at Kazutoyo Osoegawa and |
| | Cancer Genetics at the Roswell Park Cancer Institute in Buffalo, |
| | NY. The library is named RPCL-98 and was constructed by partial |
| | ECORI digestion of Drosophila DNA provided by the BDGP from the |
| | isogenic strain Y2; cn bw sp, the same strain used for the BDGP's |
| | PL and EST libraries. A more detailed description of the library |
| | and how to order individual BAC clones, the entire library, or |
| | filters for hybridization from the BACPAC Resource Center can be |
| | found at http://Bacpac.med.buffalo.edu/drosophila_bac.htm. |
| FEATURES | Location/Qualifiers |
| source | 1..1101 |
| | /organism="Drosophila melanogaster" |
| | /db_xref="taxon:7227" |
| | /clone_lib="RPCL-98" |
| | /clone="BACR08K08" |
| | /note="end : TET3" |
| BASE COUNT | 395 a 120 c 103 g 334 t 149 others |
| ORIGIN | |
| Query Match | 2.9%; Score 93.6; DB 219; Length 1101; |
| Best Local Similarity | 42.8%; Pred. No. 5.2e-06; |
| Matches 205; Conservative | 51; Mismatches 218; Indels 5; Gaps 1. |
| Oy | 1722 tgctatttttcgcgatcgatgaagaagataaacctaagtacttaatttlttgaaaatgta 1781 |
| Db | 1062 TTTTATTTTTTTTTTTTAAATATTAATTAATTTTTTTGGTGTGTTTWTWNNATTVAATTTT 1003 |
| Oy | 1782 accctttaccatacagataaattaccgcatgttttgttgcacataagcacgctctaca 1841 |
| Db | 1002 ATTTTTTTTTTTTWWATTTAATAATWAATATAAAAAATATTWTWTAATRTWTTTWARRAK 943 |
| Oy | 1842 actgtgatagtcgaatttllctgcacaataataatlaggaattcaatgctactatcata 1901 |
| Db | 942 AAAAAMAAATMTATWATTTTATTTATTTTATTTTWTWTTAMTTTAAATTTAAAAMWTAT 883 |
| Oy | 1902 gaagaacacgcgtagtatccattttaaatttaagaacaaaatttttgaaaaatgt---- 1956 |
| Db | 882 WMAATATMTATTTATTTATTTATTTTWTWTTTATTTATTAANAATWTTTAAATKAWATTTAAW 823 |
| Oy | 1957 tataatttctacaacatatatcaataalgaagcctataatgatattcctatgctctaa 2016 |
| Db | 822 TATMAAATTTTAAAMAAMTATMNNAATNATATAAATWATWATTTTTTAAATATAAATAA 763 |
| Oy | 2017 aataatttttttataattatgatacaataataacatataagaaacaataagtgiggaatc 2076 |
| Db | 762 WTATTTTATTTTAAATATTTTWTAAAMATATATATTTTAAATTTTATTTATTTTATTTT 703 |
| Oy | 2077 aaatccccataatttttgaacctacaattttaatttaattttagtgcataaacat 2136 |
| Db | 702 TATTTTTTTTTTTTTTTTTTTTTTAAATWTTTAAATATAAAMATTTATTTAAATAATTAAMA 643 |

| | | | |
|-----------------------|---|---|-----------------------------------|
| Oy | 2137 | gcataagaaagtccaaaaaatlttgttaacagaactccaatlttttlttat | 2195 |
| Dd | 642 | ANATWAAAAAAWTAAAAAATAAAAAAAAAAAAAAAAAAAATKTTTTTTTTTTTTTTT | 584 |
| RESULT . 11 | | | |
| CNSNOEO7 | | | |
| LOCUS | CNSOEO7 | 1101 bp | DNA |
| DEFINITION | Drosophila melanogaster genome survey sequence TET3 end of BAC: BACR29P01 of RpCI-98 library from Drosophila melanogaster (fruit fly), genomic survey sequence. | | |
| ACCESSION | AL069440 | | |
| VERSION | AL069440.1 GI:4949583 | | |
| KEYWORDS | GSS. | | |
| SOURCE | fruit fly. | | |
| ORGANISM | Drosophila melanogaster Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila. 1 (bases 1 to 1101) | | |
| REFERENCE | Genoscope. | | |
| AUTHORS | Direct Submission | | |
| TITLE | Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr - web : www.genoscope.cns.fr) | | |
| JOURNAL | Determination of this BAC-end sequence was carried out as part of a collaboration with the Berkeley Drosophila Genome Project (BDGP). The BDGP is constructing a physical map of the Drosophila melanogaster genome using these BACs. For further information please see http://www.fruitfly.org/The BDGP Drosophila melanogaster BAC library was prepared by Kazutoyo Osogawa and Aaron Mammosser in Pieter de Jong's laboratory in the Department of Cancer Genetics at the Roswell Park Cancer Institute in Buffalo, NY. The library is named RpCI-98 and was constructed by partial EcoRI digestion of Drosophila DNA provided by the BDGP from the isogenic strain Y2: cn bw sp, the same strain used for the BDGP's Pl and EST libraries. A more detailed description of the library and how to order individual BAC clones, the entire library, or filters for hybridization from the BACPAC Resource Center can be found at http://Bacpac.med.buffalo.edu/drosophila_bac.htm . | | |
| FEATURES | | | |
| source | 1..1101 /organism="Drosophila melanogaster" /db_xref="taxon:7227" /clone_lib="RpCI-98" /clone="BACR29P01" /note="end : TET3" | | |
| BASE COUNT | 366 a | 66 c | 104 g 351 t 214 others |
| ORIGIN | | | |
| Query Match | 2.9%; Score 93.6; | DB 219; | Length 1101; |
| Best Local Similarity | 40.4%; | Pred. NO. 5.2e-06; | |
| Matches | 246; | Conservative 83; | Mismatches 269; Indels 11; Gaps 3 |
| Oy | 1866 | aaatatgaattaggaaatcgaatgctactacaatagaagaacagcgtagattacatt | 1925 |
| Dd | 425 | ACATGAAMAAAARRGGGAGAKAATTAATTTCGAACATATAAATAAANAATTAATTAATTA | 484 |
| Oy | 1926 | ttaatttaaagcaaaaatttttgaaaaaagtataaatttcctaacaatatattataaatt | 1985 |
| Dd | 485 | TAAATWATATATATTAATAAATTTTAAAAAAMWAATAATTAATAATCAATWTAATTAATTTMT | 544 |
| Oy | 1986 | gatgcctaatgtaatttc-catgtctctaaataattttttttatcatagtataaa | 2044 |
| Dd | 545 | TATTAATTAATMAATTTTATTTATMMATATATAAANAATTAATTTTATTTATATAT | 604 |
| Oy | 2045 | tacattatagaaccaataatagttggatgaaatccaaatcctccattataatttttgaat | 2104 |
| Dd | 605 | WAAGCWMMAAATAAAMWTATAAHWTATAATTATATTAATTAATTTTAATTAATTAATTT | 664 |
| Oy | 2105 | ctacaatatataatattatgtacataaacatgcatagagaagttccaaaaaatttgc | 2164 |

[illegible]

| RESULT | 12 |
|------------|--|
| CNS00B01/c | |
| LOCUS | |
| DEFINITION | |
| | CNS00B01 1101 bp DNA GSS 04-JUN-1999 |
| | Drosophila melanogaster genome survey sequence T7 end of BAC # |
| | BACR24D09 of RPc1-98 library from Drosophila melanogaster (fruit |
| | fly), genomic survey sequence. |

| REFERENCE | ORGANISM | SOURCE | KEYWORDS | VERSION | AL057419 | GI:4937685 |
|-----------|--|------------|----------|---------|----------|------------|
| | | fruit fly, | | | | |
| | Drosophila | | | | | |
| | Drosophila melanogaster | | | | | |
| | Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta | | | | | |
| | Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; | | | | | |
| | Muscomorpha; Ephydriidae; Drosophilidae; Drosophila. | | | | | |
| | (bases 1 to 1101) | | | | | |

JOURNAL Submitted (02-JUN-1999) Genoscope - Centre National de Séquençage : BP 191 91006 Evry cedex - FRANCE (E-mail : segreff@genoscope.cns.fr - Web : www.genoscope.cns.fr)

COMMENT Determination of this BAC-end sequence was carried out as part of a

melanogaster BAC library was prepared by Kikazuho Osoegawa and Aaron Mammosser in Pieter de Jong's laboratory in the Department of Cancer Genetics at the Roswell Park Cancer Institute in Buffalo, NY. The library is named RPc1-98 and was constructed by partial EcoRI digestion of *Drosophila* DNA provided by the BDGP from the isogenic strain y2; cn bw sp, the same strain used for the BDGP's pl and EST libraries. A more detailed description of the library and how to order individual BAC clones, the entire library, or filters for hybridization from the BACPAC Resource Center can be found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.

FEATURES
SOURCE

```

BASE COUNT      420 a      56 c      48 g      261 t      316 others
ORIGIN
/organism="Drosophila melanogaster"
/db_xref="taxon:7227"
/clone_lib="RPCI-98"
/clone="BACR24D09"
/note="end : 17"

```

| | | | | |
|-----------------------|------------------|--------------------|-----------|--------------|
| Query Match | 2.9% | Score 93; | DB 219; | Length 1101; |
| Best Local Similarity | 39.2% | Pred. No. 6.5e-06; | | |
| Matches 235; | Conservative 78; | Mismatches 278; | Indels 8; | Gaps 1. |

[illegible]

| RESULT 13 | CNS020VJ/c | LOCUS | DEFINITION |
|-----------|--|-------|-------------|
| | 661 bp | DNA | |
| | | GSS | 14-MAY-2000 |
| | Tetradodon nigroviridis genome survey sequence. PUC-ori end of clone | | |

| | |
|-----------|--|
| ACCESSION | AL209800 |
| VERSION | AL209800.1 |
| KEYWORDS | GI:7868619 |
| SOURCE | GSS: genome survey sequence Tetraodon nigroviridis. |

ORGANISM
Tetraodon nigroviridis
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
Tetraodontidae; Tetraodon.
1 (bases 1 to 661)
Roest-Crolius, H., Jallion, O., Dasilva, C., Flakam, C., Fisher, C.,
Bouneau, L., Billault, A., Quetier, F., Saurin, W., Bernot, A. and
Weissenbach, J.

TITLE Characterization and repeat analysis of the compact genome of the freshwater pufferfish *Tetraodon nigroviridis*

REFERENCE

2 (bases 1 to 661)

AVTHONS
ROEST-CROLLIUS, H., JAILLON, O., BASILIVA, C., BONNEAU, L., FISH
Bernot, A., FIZAMES, C., WINCKER, P., BROTTIER, P., QUETLER, F.,

| TITLE | Human gene number estimate provided by genome wide analysis using Tetraodon nigroviridis DNA sequence |
|-------|---|
| 1 | 1 |

Db 731 RTWRRRRRRRRDMDADADDTARDDRRRRGDDGAGKGTGRRRRRRRRATWDRDTA 672

Qy 1876 ttgagatcaatgctactatcaatagaagaacagctgagatctatcaatttaattaa 1935
 Db 671 WWDAAWMTTDTDDTDMDKRRRRRRRRRTTAAAWDMWTWAWMDKMDKTRAD 612

Qy 1936 gacaaaatttggaaaatgataatctcaacaatattatataaataatgagctata 1995
 Db 611 RMDRMAADWTDARKADRWAKARAWRRARDRAARARADRWTTKGTATATWTWAAAR 552

Qy 1996 atgtattctctatgctctcaaaatattttttatataatgataaataactatga 2055
 Db 551 AAWMAWMAWTTATATWTTTWTWTTTWTWTTTAAWMAWMAWMAWMAWMAWMAWMAW 492

Qy 2056 ccaataatgctgagatcaataatccatcaatatttttgaactcaacaattat 2115
 Db 491 AAAAAAAMATTTTTTTTTTTTAAWMTAAWMTAAWMTAAWMTAAWMTAAWMTAAWMT 432

Qy 2116 taatattagtcataacatacagatagaa 2144
 Db 431 WAATATTTTWTWTTWTTAAAAAAMWMAW 403

RESULT 15

CNS04DOK 945 bp DNA GSS 21-MAY-2000
 LOCUS Tetraodon nigroviridis genome survey sequence T7 end of clone
 DEFINITION 101H21 of library G from Tetraodon nigroviridis, genomic survey
 sequence.
 ACCESSION AL285149
 VERSION AL285149.1 GI:8023560
 KEYWORDS GSS; genome survey sequence.
 SOURCE Tetraodon nigroviridis.
 ORGANISM Tetraodon nigroviridis
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
 Acanthomorpha; Acanthopterygii; Perciformes; Tetraodontiformes;
 Tetraodontidae; Tetraodon.
 1 (bases 1 to 945)
 Roest-Crolius,H., Jallion,O., Dasilva,C., Fitzmes,C., Fisher,C.,
 Bouneau,L., Billault,A., Quetier,F., Saurin,W., Bernot,A. and
 Weissenbach,J.
 Characterization and repeat analysis of the compact genome of the
 freshwater pufferfish Tetraodon nigroviridis
 Unpublished
 2 (bases 1 to 945)
 Roest-Crolius,H., Jallion,O., Dasilva,C., Bouneau,L., Fisher,C.,
 Bernot,A., Fitzmes,C., Wincker,P., Brothier,P., Quetier,F.,
 Saurin,W. and Weissenbach,J.
 Human gene number estimate provided by genome wide analysis using
 Tetraodon nigroviridis DNA sequence
 Unpublished
 3 (bases 1 to 945)
 Genoscope.
 Direct Submission
 JOURNAL Submitted (12-APR-2000) to the EMBL/GenBank/DBJ databases
 REFERENCE This sequence is a single read and was generated as part of a large
 COMMENT scale clone-end sequencing project of the Tetraodon nigroviridis
 genome. For more information, please take a look at
 http://www.genoscope.cns.fr/Tetraodon.
 FEATURES
 source
 1..945
 /organism="Tetraodon nigroviridis"
 /db_xref="taxon:99883"
 /clone="101H21"
 /clone_1ib="G"
 /note="Genoscope sequence ID : C08G101CD11Lp1-end : T7"
 BASE COUNT 386 a 112 c 96 g 231 t 120 others
 ORIGIN

Matches 248; Conservative 64; Mismatches 276; Indels 11; Gaps 3;

Qy 1816 ttgttccatgaatgaagctctacacgctgtagtcaattttctcgaatataa 1875
 Db 354 TGGAGCCAGCCAGCCAGCCGCSNNBMCTACCAATCRNNNNNNKTTTTNNHHWTTTAA 413

Qy 1876 ttgagatcaatgctactatcaatagaagaacagctgagatctatcaatttaattaa 1935
 Db 414 YANCAATTAAMNCAAMHMTTWTWTTTTTTTANTATAYMTAAWMTWTWTTTTTTTAAAT 473

Qy 1936 gacaaaatttggaaaatgataatctcaacaatattatataaataatgagctata 1995
 Db 474 TAATATAAATTTAAANWATWTTWMAAAMAAATATAAATAAAMWMTT-----TTAA 528

Qy 1996 atgtattctctatgctctcaaaatattttttatataatgataaataactatga 2055
 Db 529 AAAAAATTAATAATTTWTTAAWMAWTTTAAWMTAAWMTAAWMTAAWMTAAWMTAAW 588

Qy 2056 ccaataatgctgagatcaataatccatcaatatttttgaactcaacaattat 2115
 Db 589 ATTTTAAAAAAMATTTTAAWMTAAWMTAAWMTAAWMTAAWMTAAWMTAAWMTAAWMT 648

Qy 2116 taatattagtcataacatacagatagaaatccaaaaaatttgcacagaact 2175
 Db 649 AAWMAWMAWMTAAWMAWMAWMTAAWMTAAWMTAAWMTAAWMTAAWMTAAWMTAAWMT 706

Qy 2176 tccaaatttttttttgaagaagaataacacgataagaactatttgcgtgga 2235
 Db 707 WAAAWMAWMAWMTAAWMAWMAWMAWMAWMAWMAWMAWMAWMAWMAWMAWMAWMAW 766

Qy 2236 atggaagtagtaataacatacagaacttttaaaaaatataataacgctacgcctc 2295
 Db 767 ATAAATWMAATAAATAAAAAATTTTAAATAAWMAWMAWMTAAWMTAAWMTAAWMT 826

Qy 2296 aaagtagt-----tatctgtagtggtaataatgacatgctgagatcgaactg 2351
 Db 827 AAAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 886

Qy 2352 acaacaatgaaaaacggaatcaaatataacttaaatcaaatcaaatcaaatgagtaac 2410
 Db 887 AAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 945

Search completed: October 17, 2001, 16:25:00
 Job time: 6260 sec

Query Match 2.8%; Score 90.6; DB 221; Length 945;
 Best Local Similarity 41.4%; Pred. No. 1.6e-05;

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 17, 2001, 19:26:38 ; Search time 9769.47 Seconds
(without alignments)
6157.353 Million cell updates/sec

Title: US-09-502-426-1_COPY_3000_6888
Perfect score: 3889
Sequence: 1 atgaagctcattggttagtt.....cctagtcagccaaagctt 3889

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1344157 seqs, 7733874588 residues

Total number of hits satisfying chosen parameters: 2688314

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl:*
1: gb_da1:*
2: gb_da2:*
3: gb_da3:*
4: gb_da4:*
5: gb_da5:*
6: gb_da6:*
7: gb_da7:*
8: gb_da8:*
9: gb_da9:*
10: gb_da10:*
11: gb_da11:*
12: gb_da12:*
13: gb_da13:*
14: gb_da14:*
15: gb_da15:*
16: gb_da16:*
17: gb_da17:*
18: gb_da18:*
19: gb_da19:*
20: gb_da20:*
21: gb_da21:*
22: gb_da22:*
23: gb_da23:*
24: gb_da24:*
25: gb_da25:*
26: gb_da26:*
27: gb_da27:*
28: gb_da28:*
29: gb_da29:*
30: gb_da30:*
31: gb_da31:*
32: gb_da32:*
33: gb_da33:*
34: gb_da34:*
35: gb_da35:*
36: gb_da36:*
37: gb_da37:*
38: gb_da38:*
39: gb_da39:*
40: gb_da40:*
41: gb_da41:*
42: gb_da42:*
43: gb_da43:*

44: em_ov:*
45: em_pat:*
46: em_ph:*
47: em_pl:*
48: em_ro:*
49: em_sts:*
50: em_sy:*
51: em_un:*
52: em_vi:*
53: gb_sts1:*
54: gb_sts2:*
55: gb_sts3:*
56: gb_sy:*
57: gb_un:*
58: gb_vi:*
59: gb_v12:*
60: gb_v13:*
61: gb_v14:*
62: gb_v15:*
63: gb_v16:*
64: gb_v17:*
65: gb_v18:*
66: gb_v19:*
67: gb_v20:*
68: gb_v21:*
69: gb_v22:*
70: gb_v23:*
71: gb_v24:*
72: gb_v25:*
73: gb_v26:*
74: gb_v27:*
75: gb_v28:*
76: gb_v29:*
77: gb_v30:*
78: gb_v31:*
79: gb_v32:*
80: gb_v33:*
81: gb_v34:*
82: gb_v35:*
83: gb_v36:*
84: gb_v37:*
85: gb_v38:*
86: gb_v39:*
87: gb_v40:*
88: gb_v41:*
89: gb_v42:*
90: gb_v43:*
91: gb_v44:*
92: gb_v45:*
93: gb_v46:*
94: gb_v47:*
95: gb_v48:*
96: gb_v49:*
97: gb_v50:*
98: em_da3:*

pred. NO. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Match | Length | DB | ID | Description |
|---------------|--------|-------|--------|----|-----------|--------------------|
| 1 | 3889 | 100.0 | 4818 | 12 | AF044216 | AF044216 Arabidops |
| 2 | 3811.4 | 98.0 | 84196 | 14 | ATT3A5 | AL132979 Arabidops |
| 3 | 134 | 3.4 | 100269 | 13 | ATF18022 | AL163817 Arabidops |
| 4 | 128 | 3.3 | 106142 | 12 | AC025417 | AC025417 Genomic S |
| 5 | 94.8 | 2.4 | 90459 | 12 | AC012396 | AC012396 Arabidops |
| 6 | 93.2 | 2.4 | 87835 | 12 | AB005237 | AB005237 Arabidops |
| 7 | 92.8 | 2.4 | 4937 | 9 | AR074364 | AR074364 Sequence |
| 8 | 92.8 | 2.4 | 4937 | 13 | ATCYP450D | X87368 A.thaliana |

[illegible]

| | | | |
|----|------|---|------|
| QY | 1621 | atacttttgaaggagaagcattcttgagacgcgagggccagaagaagcaactgaggagatca | 1680 |
| Db | 2550 | ATACTTTTGAAGGAAAGGCACTCTTTGGATCGCGAGGCGCCAAAGAGCAACTGAGAGAGTCA | 2609 |
| QY | 1681 | gaatlaaatbtagagatgttacaagaagaatggacattaccgaatgtatattacatcat | 1740 |
| Db | 2610 | GAATTAATTTGGATGTATTCCAAAGAAAATGGACTTTACTCAATGTGTATTACTATCAT | 2669 |
| QY | 1741 | tctaatattatattcattgtctatcatatgtatcatatgaataccaatatattgatttttt | 1800 |
| Db | 2670 | TCTCATTTATTATTCTATGTTCATATGATTTATGATGAACCAAAATTTATGATTTTTTT | 2729 |
| QY | 1801 | tttgatgtgtgtgaaggcttaaatgaaaccttcgatttgggaaatgtatgtagttttt | 1860 |
| Db | 2730 | TTTGGTGTGTGTGAAGGTTTAAATGAACCTTCATTTGGGAAATGTATGATGTTTTT | 2789 |
| QY | 1861 | gcatacgaaagcaactcaagaagatgttcggtgacaagaagttaaactttacgtacaaaatttt | 1920 |
| Db | 2780 | GCATCGCAAGCAACTCAAGAGTGTGGTACAAAGGTAAACCTTTACGTCAAAATTTTTT | 2849 |
| QY | 1921 | aaataatgaaaatccgugaatatitgaactcttatitgagatgaanaaataltaaatlaattica | 1980 |
| Db | 2850 | AAATATATGAATCCGGATATFTTGAATCTTATTTGGATGAANAATFTTAAATATTTTCA | 2909 |
| QY | 1961 | tttctttaatgtctgaaaaaagagataacctatccctagtggtgtgagaagtgltaccggtga | 2040 |
| Db | 2910 | TTTTCTATATGTTGGAAAAAGATACGATATCCCTAGTGGGTGAAGAGTTTACCGGTGA | 2969 |
| QY | 2041 | tctcaagccgtacatttbgatataatctcgttatgacaaactaatctctttaaactcttgga | 2100 |
| Db | 2970 | TCTCAGCCGTACATTTTGGATTAATTTCTGTTATACCAACCTTATCTCTTTAATCCTTGA | 3029 |
| QY | 2101 | gatggcaacaggtlaaataaaaagtttctcgtctaaactacgaaatagtgatagattt | 2160 |
| Db | 3030 | GATGGCAACAGGTAAATTAAGATTTTCTGTTACTTATCTATCTATTAAGTATAGTTTT | 3089 |
| QY | 2161 | tttaactatctgaatgaatagataagtcactgaagtgatttaactatcatagatactataac | 2220 |
| Db | 3090 | TTTTCACTATTTGCATGTAGTATGATACGCTACACGTGATTTACCTATCTATTAAGTATATAC | 3149 |
| QY | 2221 | gagaactatgaactgacaaaacttttcttatttcttcacagttaactcttaac | 2280 |
| Db | 3150 | GAGAACTATTAAATCTGGCAAAACTTTTATTATTATTCTTTCAAGTTAGATCTTTAAC | 3209 |
| QY | 2281 | acgtcatgagatcatgtgacatgatgaagaacatataaaatlaaagaagagagaagaaga | 2340 |
| Db | 3210 | ACGTCATGATCATTTGATACATGAAGATTAATTAATTAATAATTAAGAGAGAAAGAGA | 3269 |
| QY | 2341 | cgtgtgtgtgtgaagtgtaacgtgaaacaatlaattagtagagtgatgtactttaaagac | 2400 |
| Db | 3270 | CGTGTGTGTGTAACTGTATGATGAGAAACAATTAATTTAGTATGATGCTTTTAATATAC | 3329 |
| QY | 2401 | gttagagagctgcctaataattccttataatctgtgacggttgatattatattagccaagctt | 2460 |
| Db | 3330 | GTAGAGCTGCTCTAAATATTTCTTATATCTGTGACCGCTGATTTATTTATTTAGTCACGGCTT | 3389 |
| QY | 2461 | tgaatacaatttaagaatttgacgcgacgtgttacaacgagcttgaacgagatctacaacgccc | 2520 |
| Db | 3390 | TGATATCAATTTAAATATTGAGAGCAATGATGATACAGGCTTTGACGAGATCTCACAGGCC | 3449 |
| QY | 2521 | gatgactgtacgtgcgttagatctgcgcacglttgacggttttaactatagatttata | 2580 |
| Db | 3450 | GATGACTGTACGTCGCTGTAGATTTCTGCCACGTTGACTGCTGTTTAAATCTTTGATTTATA | 3509 |
| QY | 2581 | actcatattaactataaactatcaaatcggcggaatttagaagaanaatatactataagttt | 2640 |
| Db | 3510 | ACTCATTAATTTAATACACTATCAAAATCCGCAATTTAAGAAATTAATCTATTAATGTATT | 3569 |
| QY | 2641 | attatgattatcatgagataaactatcatataaataagatcaatlaaivgttagatcatgtt | 2700 |
| Db | 3570 | ATTATAGATTATTATAGACATTAATCTTTATGCAAAATTAAGATTAATTAATGATCATATGTTT | 3629 |
| QY | 2701 | atagtgatgtgggaaglaaagagtggtgtgagagatgattaatgacccacgltgtgtgtgt | 2760 |

[illegible]

| DB | 4710 | ATCAGCTTTAGAGATCCTCTTCTCAGACAGTAAGTAATCCCTGCTTACCAAGCAAGATGCT | 4769 |
|--------------|--|---|-------------|
| QY | 3841 | catctttgaagacaggtcttccaaagtcctagttcagggccaagctt | 3889 |
| Db | 4770 | CATCTTTTGAAAGACAGGCTCTTTCCAGTCTTCAAGTCTTCAAGGCCAAGCTT | 4818 |
| RESULT | 2 | | |
| ATT3A5/c | | | |
| LOCUS | ATT3A5 | 84196 bp | PLN |
| DEFINITION | Arabidopsis thaliana DNA chromosome 3, BAC clone T3A5. | | 26-JAN-2000 |
| ACCESSION | AL132979 | | |
| VERSION | AL132979.2 | GI:6782244 | |
| KEYWORDS | | | |
| SOURCE | | | |
| ORGANISM | thale cress. | | |
| REFERENCE | Arabidopsis thaliana | | |
| AUTHORS | Eukaryota: Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis. | | |
| REFERENCE | 1 (bases 1 to 84196) | | |
| AUTHORS | Bloeker,H., Mewes,H.W., Lemcke,K., Mayer,K.F.X., Quetier,F. and Salanoubat,M. | | |
| JOURNAL | Unpublished | | |
| REFERENCE | 2 (bases 1 to 84196) | | |
| AUTHORS | EU Arabidopsis sequencing project. | | |
| TITLE | Direct Submission | | |
| JOURNAL | Submitted (25-JAN-2000) MIPS, at the Max-planck-institut fuer Biochemie, Am Klopferspitz 1ba, D-82152 Martinsried, FRG, E-mail: lemcke@mps.biochem.mpg.de, mayer@mps.biochem.mpg.de, Project Coordinator: Marcel Salanoubat and Francis Quetier, Groupement d'Interet Public, Centre National de Sequencage - GENOSCOPE; 2 rue Gaston Cremieux, BP191, 91006 Evry Cedex, France; http://www.genoscope.cns.fr | | |
| COMMENT | On Jan 27, 2000 this sequence version replaced gi:6424247. Information on performance of analysis and a more detailed annotation of this entry and other sequences of chromosomes 3, 4 and 5 can be viewed at: http://www.mips.biochem.mpg.de/proj/thal/. | | |
| FEATURES | | | |
| source | location/Qualifiers | | |
| | 1..84196 | | |
| | /organism="Arabidopsis thaliana" | | |
| | /variety="Columbia" | | |
| | /db_xref="taxon:3702" | | |
| | /chromosome="3" | | |
| LTR | 2983..3338 | | |
| | /note="355 bp LTR" | | |
| misc_feature | 3339..7564 | | |
| | /note="LTR-transposon" | | |
| LTR | 7565..7922 | | |
| | /note="367bp LTR" | | |
| gene | 12339..13197 | | |
| | /gene="T3A5.10" | | |
| CDS | join(12339..12665,12742..12803,12888..13055,13125..13197) | | |
| | /gene="T3A5.10" | | |
| | /note="similarity to cyclin-dependent kinase inhibitor (IC1), Arabidopsis thaliana, EMBL:AF094772.1" | | |
| | /codon_start=1 | | |
| | /product="putative protein" | | |
| | /protein_id="CAB62432.1" | | |
| | /db_xref="GI:6561966" | | |
| | /translation="MAAVRRRRRDVVEENGVTITTVKRRKMEEEVDLVESRIISPCVQATRRGIVARNSAGSETSVIYRRDSPVEEOCIIEEDSSVSCSTSESKRRLEFDLENNGDRETTSMYIDLKNSSESMNDDSSVADEVESRRIRKLSHETVKEALDEPFOVAEKDLRNKLLLECSMKNPDEKDEPLGGRTYEWKLP" | | |
| exon | 12339..12665 | | |
| | /gene="T3A5.10" | | |
| | /number=1 | | |
| intron | 12666..12741 | | |
| | /gene="T3A5.10" | | |
| | /number=1 | | |
| exon | 12742..12803 | | |
| | /gene="T3A5.10" | | |
| | /number=2 | | |
| intron | 12804..12887 | | |

/gene="T3A5.10"
 /number=2
 12888. .13055
 /gene="T3A5.10"
 /number=3
 13056. .13124
 /gene="T3A5.10"
 /number=3
 13125. .13197
 /gene="T3A5.10"
 /number=4
 complement(15858. .15901)
 /number=1
 15858. .16470
 /gene="T3A5.20"
 complement(join(15858. .15901,16014. .16470))
 /note="Similarity to predicted proteins, Arabidopsis thaliana"
 /codon_start=1
 /product="putative protein"
 /protein_id="CAB62433.1"
 /db_xref="GI:6561967"
 /translation="MARLEHNSMNGARISFSENFVEIRSEKSNKSNINNSRSES
 MPSADFAFVTDYSMIPADEIFLGKILPEKTSVHRTIGEELLTEEGSMVGNTF
 SLRPLSSSESTKGTWRELLGLKRTHVRSKTDKVNVEVLSQDHKIISGVATREC
 OVADTR"
 complement(15902. .16013)
 /number=1
 complement(16014. .16470)
 /number=2
 18303. .19931
 /gene="T3A5.30"
 complement(18303. .19931)
 /number=1
 complement(18303. .19931)
 /codon_start=1
 /product="scarecrow-like 7 (SCR7)"
 /protein_id="CAB62434.1"
 /db_xref="GI:6561968"
 /translation="MAYVCTDSGNIMATAOOLIKOKOOOQOEOEOEOPNPMN
 PSFGTLPGSGFSDPFOVNDPGRPHLEHQNAAVASEEDSDMESLNGSDAS
 QTNDFPIYGHDPFVSPSRLSAPYLPHNKKDDSAQOLPPPPASTAWISPSPPQ
 HPPPPQPPDLDLNPDKAIHDYARKPETDILIRIKESVSESGDIOORGVFAE
 ALSHETSPSSSSSLEDRLISYKTLNDACPYKFAHANOAILLATONNNIHI
 VDFGFGIOGSAIIOALATRSSGKPTIRISGIPASIGSPGSLATONRLEDA
 AILDNEFFPYPLPIQLNGSSFRVDDEVLYVFMLELYKLDDETATYGTALRLA
 RSLNRIYLTGEYSILNRVEFANRKNLSLFYSAVFESLEPNLDRDKEKRLRERVL
 FGRIMDLVRSDDDNNKPGTRFGIMEEKEOMRVLMKAGFEPVPSNYAVASQAKLLM
 NYNSTLSVSESGFISLANNVPLTVSSWR"
 26093. .28999
 /gene="T3A5.40"
 complement(join(26093. .26338,27091. .27200,27308. .27386,
 27478. .27570,27723. .28037,28136. .28288,28374. .28698,
 28779. .28999))
 /note="Contains Cytochrome P450 cysteine heme-iron ligand
 signature AA455-464; Prokaryotic membrane lipoprotein lipid
 attachment site AA320-330"
 /codon_start=1
 /product="steroid 22-alpha-hydroxylase (DMF4)"
 /protein_id="CAB62435.1"
 /db_xref="GI:6561969"
 /translation="MFEETHHTLPLLLPSLISLFLILIKRRNRKTRFLPPGKS
 GMPLEGTICGLKPYATTGLDPMQOHVSKYKIRSNLSEPTIYSADAKNRIILO
 NEGRLECSYRSISIGILGKMSMLVGDMDHMSISLNFISHARKLTILKIDVERH
 TLFVLDTSWQNSIFSAODEAKRFTFNMAKHMSMDPEEETQAKKYVFMKQVSA
 APLNPGTAAYKALOSRAITILKTERKMEERLIDKEDEEVEYTKDEDEMSKVS
 VRKORTDLDLGMVKHNSLSTEOILDLISLDFAGHESVAIALAIFLQACAKAV
 EELREHLEIARAKKEIGESLNMDDYKMKMTOCVINMTLRIGVAVFLFKKADKY
 RYKGYDISGKVKVLPYISAVHLDNSRYQPNLFNWRMOONNGNASSSGSGSFISWGN
 NYMPGGGPRILCAGSELAKLEMAVFIHLLVLKFMNLEADDPFAFPVDPENGLPIR
 VSRIL"
 complement(26093. .26338)
 /number=1

intron
 /complement(26339. .27090)
 /number=1
 complement(27091. .27200)
 /number=2
 complement(27201. .27307)
 /number=2
 complement(27308. .27386)
 /number=3
 complement(27387. .27477)
 /number=3
 complement(27478. .27570)
 /number=4
 complement(27571. .27722)
 /number=4
 complement(27723. .28037)
 /number=5
 complement(28038. .28135)
 /number=5
 complement(28136. .28288)
 /number=6
 complement(28289. .28373)
 /number=6
 complement(28374. .28698)
 /number=7
 complement(28699. .28778)
 /number=7
 complement(28779. .28999)
 /number=8
 38307. .41323
 /gene="T3A5.50"
 complement(38307. .38796)
 /number=1
 complement(join(38307. .38796,38877. .39090,39189. .39270,
 40180. .40242,40337. .40401,40489. .40564,40655. .40734,
 40805. .40919,41225. .41323))
 /note="Contains Eukaryotic putative RNA-binding region
 RNP-1 signature AA179-186"
 /codon_start=1
 /product="U1 snRNP 70K protein"
 /protein_id="CAB62436.1"
 /db_xref="GI:6561970"
 /translation="MGDSGDPLRNPNNAVOARAKVQNRANVLOLKLKGSHPTGLTN
 NLKLFEEPRPLEYKPPPEKRCQPYTGAQFVSFAPGDEYAPPEVEYLPQKR
 ERIHRLTEKGKEKAEDLKRYDPNNDNATGDPYKTLFVSLNTESESKIKREES
 YGPIKRYVLVTDQLNKRKGAFTYHMTROMKAAVKKADOCKIDGRVYLDVNGRT
 VPMNPRRLGGGLGSRVGGGEIVGEQPGRTSOSSEPSRPERKSRKGGERE
 SRSLSHQDPKRSRDREREDKHHNRDQGGRRDRDRDRDRTRDGDRDRDR
 GDRTSRHDHDSRKKEEDYEGGEYEHGGGRRSERDAEYRGEPEETRGYYEDOGD
 TDYSHRYDKMEDDEFRYERKRSRESREYVR"
 complement(38797. .38876)
 /number=1
 complement(38877. .39090)
 /number=2
 complement(39091. .39188)
 /number=2
 complement(39189. .39270)
 /number=3
 complement(39271. .40179)
 /number=3
 complement(40180. .40242)
 /number=4
 complement(40243. .40336)
 /number=4

Query Match 98.0%; Score 3811.4; DB 14; Length 84196;
 Best Local Similarity 99.7%; Pred. No. 0;
 Matches 3881; Conservative 0; Mismatches 6; Indels 6; Gaps 6;

QY 1 atgaagcattatggtta-gtttaagcattataagaagatttcaatataatgaagaa 59
 |||||
 Db 29203 ATGAAGCATTATGGTTAAGCTTAATAAGAGATTATTATTAATAATGACGA 29144
 QY 60 tgataacaattatattcttgactctttaaaaccccccttaacaaacgaagctccct 119

|||||
Db 29143 TCATAACATTTATTTCTGACTCTTTAAACCCTCTTACAACAGAGCTCCCTT 29084
OY 120 ttctagtagaagtcgcatcccaatctttaagaacaagccatltagaagaagaatgagat 179
Db 29083 TTTGCTTAAGATCCGATTCCTCAATCTTAAAGACAAAGCCATTAGAAAGAGAGT 29024
OY 180 gagagagagagaaactagctccatglttcgaacaagacatcatctcttaccctt 239
Db 29023 GAGAGAGAGAGAAACTAGCTCCATGTTCCGAACAGACATCATCTCTTACTCTT 28964
OY 240 ctctctcccatcgcttctctctctctctctctctctctctctctctctctctct 299
Db 28963 CTTCCTCTCCATCCGCTTTGCTCTTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 28904
OY 300 agaaaaacacagatcaatctacctccggttaatccggttgccatctcttgtaaacc 359
Db 28903 AAAAAAACCATTCATCTACCTCCGGTAATCCGGTTGGCCATCTCTTGTAACC 28844
OY 360 atcggltatcttaaacggtacacgcgcaacaacacgcgtgactcatgcaacaacatg 419
Db 28843 ATCGGTATCTTAATACCGTACACGCGCAACACCTCGGTGATTCATGCAACACATGTC 28784
OY 420 tccaaagtaacaacaacatctccaaaaaactcaaaaaataaaccctctglttgaat 479
Db 28783 TCCAAAGTAAACAACAACATCTTCCAAAACTCAAAAAATAATCTCTGTTTGAAT 28724
OY 480 ttgactaaatggttctattttaacagtatgtaagatatatagatcgcgaactgttgag 539
Db 28723 TTGACTAATGTTGTTATTTTACAGGTATGTAAGTATATAGATGAACCTGTTGGAG 28664
OY 540 aaccaacgactcgtalcacgcgtga tgcgtgactcaatagatcatatcaaaaaagagaa 599
Db 28663 AACCAACGATCGTATACAGCTGATGCTGGACTTAATAGATTATTAACAAAGAGAA 28604
OY 600 ggcctcttgaaatgtagtataccttagaagaatagtgaggatctcttgagaaatgctcga 659
Db 28603 GGCCTCTTGAATGTAGTATCTTAAAGATAGTGGGATCTTGGAATATGTCGATGC 28544
OY 660 ttgtctctgttgtagacatgcatagatatgagaatgctcgcgttaactcttaagtc 719
Db 28543 TTGTTCTTCTGTTGATGATGATAGATATGAGAAATGTCGCTTAACCTTTAAGTC 28484
OY 720 acgcacgtcttagaactcttactcttaagaatglttgagagacatacttgttctctg 779
Db 28483 ACGCAGCTTCTAGACATATTCTACTTAAAGATGTGAGACATCTTGTGTTCTTG 28424
OY 780 attcttgcaaaaaaactctatcttctcgtccaagaagcgccaaaaagtttctatt 839
Db 28423 ATTCTTGGCAACAAACCTTATTTCTGCTCAAGACGAGGCCAAAAAGGTTTATTT 28364
OY 840 ttaactcttattctgtaaatcttctgtttaagaactcttgagtttcttaactcttct 899
Db 28363 TTAATCTTTATTTTCTGCTAAATTTTGTGTTATGAATCTTTGAACTTTCTTAACCTTTT 28304
OY 900 ttcttaattgaacaglttaaglttaactcaatgycgaagcatataatgaatltgagctc 959
Db 28303 TTTTAAATTAAGACACTTAACTTAACTTAATGCGCAAGCATATATGATGATCCT 28244
OY 960 ggaagaagaagaacagagcaatltaaagaagaatgtatgtaactcttaagaagaatgtct 1019
Db 28243 GGAGAGAGAGAAACAGACCAATTAAGAAAGATGTGTAACCTTATGAAAGACTTGTGC 28184
OY 1020 tctgctcccttaactcccaagaaactgttactcaataagctctcaggtacatctatt 1079
Db 28183 TCTGCTCCCTTAATCTACAGAGAACTGCTTATCAATAAGCTCTTCAGGTACATTTATTT 28124
OY 1080 ttcttctgtaagaatgcaacaactctcaatlataggttttaattttaattgtaagta 1139
Db 28123 TTTTGTGCTGAAGTCAACAACTCTCATTTATAGGTTTATTTATTTATGTTAA 28064
OY 1140 ataaatatctaaatggttgtagtcaagagacatatgtgaagttcattctgagaga 1199
|||||

Db 28063 ATAAATATCTAAATGCTGTGTAGTCAGAGCAACATATTGAAGTTTATGAGAGA 28004
OY 1200 aaatggaagaagaataattgga latcaaggaagaagatccaagaagaagaagaa 1259
Db 28003 AAATGGAAGAGAAATTTGATATCAAGGAAGATATCAAGAAAGAGAAAGAGAA 27944
OY 1260 cagaagatgaagcagaagatgaagtaagatgtaacatggttaagaacaagaacagacag 1319
Db 27943 CAGAGATGAAGACAGATGATGATTAAGATGATCATGTTAGGAACAAAGAACAGACGATG 27884
OY 1320 atcttctggagatgggttttgaacaatcgaatctatcgacggagcaatctcgaatc 1379
Db 27883 ATCTTTGGGATGGGTTTGAACATTCGAATTTATCAGGAGCAAAATTCGATCTCA 27824
OY 1380 ttcttaagttgttatttccggacatggaactctctctgtagccatgctcgcgtatct 1439
Db 27823 TTCTTAGTTTGTATTTTGCCGACATGAGACTTCTTGTGACCATGCTCTCGCTATCT 27764
OY 1440 tctcttgcaagcttgccctaaagccgttgaagagcttaggtaagaataatataacagc 1499
Db 27763 TCTTCTGCAACCTTGCCCTAAAGCGTTGAAGACCTTAGGTATGATATTAACAGC 27704
OY 1500 acaagttaacttaaccaaatgttagtaltatataagttatataagatlatatctatc 1559
Db 27703 ACACTTAATTAATCTCAAAATTTGATCTATATTAATTAAGTTTATTAAGATTAATCTAT 27644
OY 1560 agaataacgaatgaaaaaagatgafatatttaattgtcaactatttgattatgatt 1619
Db 27643 AGAATATTCGATGAAAGAAAGTATGATATTTATTAATGTCATTAATTTATGTTATGAT 27584
OY 1620 tatacttctgaagaagacatcttgaaatcgcgagagcgccaaagaagaacatagagaatc 1679
Db 27583 TATACTTTGAAGAGAAAGACATCTTGAGATCGCGAGGCCAAAGAAAGAAATAGAGAGTC 27524
OY 1680 agaataaattgaggatgattatacaaaaaaagacttaactcaatgtagtactactaca 1739
Db 27523 AGAATTAATTTGGATGATTTCAAGAAAGATGAGACTTCTCAATGTTATGTTACTATCA 27464
OY 1740 ttctcatattatctatctatgltcatatgatltatgataaacaacaactatagactctt 1799
Db 27463 TTCTCATTTATTTATTTCTATGTTTATGATGATTTATGATGAAACAAATTTATTTATTT 27404
OY 1800 tttagtgltgltgaagltataaaatgaactctcgaatcttggaagatglttagttc 1859
Db 27403 TTTTGTGTGTCTGAAGTTTAAATGAACCTTCCGATTTGGGAATGTAGTTAGCTTT 27344
OY 1860 tgcacgcgaagaactcaagaatgltcgttacaaaggtlaaaacttcaatcaaaatctt 1919
Db 27343 TGCATGCAAAAGCACTCAAAAGATGTTCCGTACAAAGGTAAACCTTACGTCAAAATTTT 27284
OY 1920 taaataatgaaaatccggaatatltgaactctatltgaatgaaaatattcaaaatattac 1979
Db 27283 TAAATTAATGAATAATCCGGAATATGAAATCTTATTTGATGAAATAATTAATTTAT 27224
OY 1980 atttcttaatglttg -aaaaaagatagatalccctagtggttggaaggtgtacgcgc 2038
Db 27223 ATTCTTAATGTTGGAATAAAAGATAGATATCCCTGAGGTGGAAGGTGTACCGGT 27164
OY 2039 gatctcagccgttaacttggatlaattcgtltatgaccaactcaatctcttaactctg 2098
Db 27163 GATCTCACCCGCTCAATTTGATATTCGTTATGCCAACCTTAATTCCTTTAAACCTTTG 27104
OY 2099 gagatgcaacaggttaaaataaagttctcctcgttactactcgaaaatagtgtagt 2158
Db 27103 GAGATGCAACAGATTAATTAATAAGTTCTCTGTTAACTATCGAAATTTAGTGTATGT 27044
OY 2159 ttcttcatctatlgcatgaatagatagctcctacgttgatlttaactcatatagatactac 2218
Db 27043 TTTTTCATCTATTGCAATGAATAGATAGCTCTACGTGATTTACATATATAGATATAT 26984
OY 2219 acggaacatataatctggcaaaaacttctatatactatcttctcaagttagatctca 2278
Db 26983 ACGAGAACTAATTAATCTGGCAAAACCTTTTATTAATTAATCTTTCAAGTTAGATCTTA 26924
|||||

[illegible]

| QY | 3359 | cttaaatgggcttctctataaggcccaatataatcagcttataaccaagtgacaacttt | 3418 |
|------------|--|---|-----------------|
| QY | 3359 | cttaaatgggcttctctataaggcccaatataatcagcttataaccaagtgacaacttt | 3418 |
| Db | 25845 | CTTAATATGGGCTTCTCTAATAAGGCCCAATTAATTAATTCGATTAATAACAAAGTGCAACTTTT | 25786 |
| QY | 3419 | acttcgtttctgacccgaagcaataaacaattgtcacaatataccaacacacaagaattatgta | 3478 |
| Db | 25785 | ACCTCGTTTTCATCCGAAGCAATTAACAATTTGTCMAATTAACCAACACAGAAATTAATGTA | 25726 |
| QY | 3479 | aacactcgttgctgtctcagtgaggaaatcaattgagcttgagactgaacatcagaacacaag | 3538 |
| Db | 25725 | AACACTCGTGCTGTCTAGTGGGAATCATCTMGCGCTGGAGACTGAAACATACGAMCACAAG | 25666 |
| QY | 3539 | aaacctgtaatatagatagataccctccatagacggtttccaaacttatacttgatctcta | 3598 |
| Db | 25665 | AAACCTGTCAATTATAGGATACACCTCTATGACGGCTTTCCAAACTTTTATCTTGATTTCTTA | 25606 |
| QY | 3599 | tgcgtttacattgacacacaagaagtagtgytcaaaaagacataatgaataacaatagctc | 3658 |
| Db | 25605 | TGCTGTTCATTTGACACAAAGAGTTTAGTGTCATAAAGAGCTAAATGAATTAACATAGCTC | 25546 |
| QY | 3659 | taagataagaagagcttcataaataatggcttcttattttgagaagaagagagagactt | 3718 |
| Db | 25545 | TCAGAGTATAGAGAGTTCATAAATAGTTTCTTTATTTTGAGAAGAAGAAGAGAGAGAGCTT | 25486 |
| QY | 3719 | ttacgtgttcttggtgtccatctcctttaaataagaagggtttgcttttactctctctc | 3778 |
| Db | 25485 | TTACGTGTTTCTTGGGGCTCTATTCCTTTTAATATAGAGGGTTTCTGTTTTTACTCTCTATAC | 25426 |
| QY | 3779 | tcatcactcttagagatccctctctagacagagaaagtaatccctgttacaagaacat-9g | 3837 |
| Db | 25425 | TCATCATCTTTTAGAGATCCCTCTCTAGACAGATTAAGTAATCCTCGTTACCAAGCAATGGG | 25366 |
| QY | 3838 | tctcatctctttaaagaca-9gtcctttccaagtccctagcttcagcgccaaagctt 3889 | |
| Db | 25365 | TCCTCCTCTTTGAAGACAGAGGCTTTTCCAAAGCTTATGTTGAGGCCAAAGCTT 25313 | |
| RESULT | 3 | | |
| LOCUS | ATF18022 | 100269 bp | PLN 13-APR-2000 |
| DEFINITION | Arabidopsis thaliana DNA chromosome 5, BAC clone F18022 (ESSA project). | | |
| ACCESSION | AL163817 | | |
| VERSION | AL163817.1 | GI:7573446 | |
| KEYWORDS | thale cress. | | |
| SOURCE | Arabidopsis thaliana | | |
| ORGANISM | Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsids. | | |
| REFERENCE | 1 (bases 1 to 100269) | | |
| AUTHORS | Bevan,M., Murphy,G., Ridley,P., Hudson,S., Bancroft,I., Mewes,H.W., Ruddle,S., Lemcke,K. and Mayer,K.F.X. | | |
| JOURNAL | Unpublished | | |
| REFERENCE | 2 (bases 1 to 100269) | | |
| AUTHORS | EU Arabidopsis sequencing project. | | |
| TITLE | Direct Submission | | |
| JOURNAL | Submitted (13-APR-2000) MIPS, at the Max-Planck-Institut fuer Biochemie, Am Klopferspitz 18a, D-82152 Martinsried, FRG, E-mail: lemckemips.biochem.mpg.de, mayer@mips.biochem.mpg.de, Project Coordinator: Mike Bevan, Molecular Genetics Department, Cambridge Laboratory, John Innes Centre, Colney Lane, NR4 7UD Norwich, UK, E-mail: michael.bevan@bbsrc.ac.uk | | |
| COMMENT | Information on performance of analysis and a more detailed annotation of this entry and other sequences of chromosomes 3, 4 and 5 can be viewed at: http://www.mips.biochem.mpg.de/proj/thal/ . | | |
| FEATURES | | | |
| source | Location/Qualifiers | | |
| | 1..100269 | | |
| | /organism="Arabidopsis thaliana" | | |
| | /variety="Columbia" | | |
| | /db_xref="taxon:3702" | | |
| | /chromosome="5" | | |
| gene | 2255..6118 | | |

CDS
 join(1701..1663,1982..2194,2289..2428,2519..2614,
 2962..3103,3181..3226,3339..3483,3599..3742,3849..3968,
 4112..4168,4249..4308,4456..4563,4722..4864,5060..5138,
 5247..5309,5402..5532,5618..5723,5908..6096,6220..6495,
 6595..6750,6873..7034,7168..7284,7367..7543)
 /note="unknown protein; similar to ESTs gb|A1999041.1,
 and dbj|AV547742.1"
 /codon_start=1
 /evidence=not_experimental
 /product="t12c24..2"
 /protein_id="AAF88074.1"
 /db_xref="GI:9502367"
 /translation="WDGRQVPSVDLERYATKNRMITCAAGNDVLTCTSGWII
 RYDFGVSSNDIDLAVNRGTGEOSIHKYFVDPGSGCATVVGGAETFPYHAKMLP
 RLSRGLKLVNAAMNRQITTEVSTKEIILTGQDLFEVADGKREYIKFLE
 LEELEAPKALOMETANISGMRYYAVATPTLRYSPFTIGLSEFASYSKREYHFM
 ELGPEIPRQVLYDLRLIRIKETIKNSTLTITWTGYHGILNGFAOHSYNGENY
 ENKALDYSKSDGTEAVKPSMSALSEHPLILGNKKYVNRISQIIEELDITIS
 DVSRRGITGLCSASANVFATDOSTFQVSVIDGRMRFVLDLKYAALALNCRD
 PLQRQVYLQVESAFTDKEXYLRASFPYKINVISEEVTLKISINEPALRFL
 HKLNLSDKDCQITIMSTWATELYLDRINRLLEDATAENRSEHSVIOEPFAEM
 SDCKDELDEATVTKLESYGVEELVYFANLKEOYEIVLHYIOGGEAKLVELOKS
 SVSEIYOFAPELIMLDAVEFVSEWMANKNLNPRRLITAMMRYSSGPHAKNETHVI
 KYLECVHRLNEDPGIHSLSITAKOEDGALLRFLQCFKRGRENGPEFFDPRK
 ALRLCKERRRACVHITSMSMEHVALALQCAESLVDLKTVSVLIQIDPELMAAE
 ADKVEDDEDLRRKRLMVAKHVQKQKAKRENIKRIATFLKTDGLKIDIDLEFPF
 DFALIDPFEKACISLEDYDKOIEOLKEENDATRGADININDISALTQRYAVIDRDE
 ECGVKRKILMMSGDFRMAOGSSAGPLAPFYVPGHSGFHAOCILITHVSCAHEOA
 EHILDLKOLLGSEMPRDINGNSDEPTSTTADKRSLELDAISSECPFGCELM
 INETLFPTRKEDSQYSTWDRSETNLANORTISLVP"
 complement(join(7769..8120,8198..9009,9085..9591))
 /note="hypothetical protein"
 /codon_start=1
 /evidence=not_experimental
 /product="t12c24..3"
 /protein_id="AAF88102.1"
 /db_xref="GI:9502395"
 /translation="MERKQNSAHSTFADINEVEDAEQELOOEENNNKRFSGNCRPN
 RGKORPFRGROVSLSTGFSVLNRESERDDKSLPRSGRSGFGESGIIINGDGR
 KTDSPMRTKSTLSOKSLPSIIRERDIENSLTRDEGTEDKDSINENVASGRYFAAL
 RGPDELVEKDNEDILLPKREOMPLRPIRGFCIGLSSQVNLMLALASPRATNPL
 HITPLINLVWLFSLVLYSVSFTYLLCITYFEAVKREYHPVRVNFEPAPWVCMF
 LAISVPMESPNRRYLHPALWCVMGPYFLELKIYQGLSGKRLCKVKNPSHLS
 VVGNEVGAILASKVGMDEAVKFLMAVFAHLYVLTQRLPTSEALPKELHYP
 FIAAPASAIAMNTIYGOFGCSRCEFIALFELISLVARINFTGFEFSAVMSYTF
 PMTASVATIKYEAVEVPGSRALATLSTFISTAMVCLFEVSTLLHAVMQFTLPNDL
 AIAITKRLTREKPEPKKAVDLKRWTKAKLSAEKDFAEBSHH"
 join(11098..11279,11308..12211)
 /note="hypothetical protein"
 /codon_start=1
 /evidence=not_experimental
 /product="t12c24..4"
 /protein_id="AAF88075.1"
 /db_xref="GI:9502368"
 /translation="MSKORCLRMESLPHEVERILKTKCGFSAELHGCITMETIID
 RISTLPKTDITASAIRSEDEYFNPDLKHLTLVLGSSSVTITPPEKDEKEREKE
 KEDEEFLVSEDCDGLCYLYKWKSGVYVTRKWPRLDSLOQOLILGISVVE
 LGTYVCDIGFCKDKITGYKPYWMLYSLIELENAATCEVDFMTNARVYSPAPR
 EETKILSPDLHETEPFVAVSKAPTNVKAFODIVMNLGNRLCVSKKNPNQVINSFNG
 NKTWHKMSINLDTSHHFGNHIAVWPLAFYKKKKKKLLFCYRVRSLTMYDPE
 TESTYVAFNDSTIGPLCYFQSLISIS"
 complement(join(13884..14297,14628..15299))
 /note="similar to phosphate(phosphoenolpyruvate
 translocator precursor gb|AAB0467.1; similar to ESTs
 dbj|AV524415.1, gb|A1995339.1, and dbj|AV542831.1"
 /codon_start=1
 /evidence=not_experimental
 /product="t12c24..5"
 /protein_id="AAF88101.1"
 /db_xref="GI:9502394"
 /translation="MVEAQSWTTRMSNPRDAAATAAPTIVDIPGTPHSSASSPLK
 PFTLSPVSPFTIILAAITAAWGSNIGVLLNLYLFFYGEFRPIELTMTMHMSCAA
 YSSAVINIAIGVPROHILSRQFLKILSLAISFCLSVVCGNTSLRYIPVSPNOIGAT

CDS
 TPFTAVFSFLITCKTESTEYVLLALPVVSGIVLASNESEPHLFGFLICVASTAGRA
 LKSVQGIITLSESEKLSHNMULLYMAPMACILLPFLYLEGVNLRVLIERKATDPL
 ILFLAGNAVAYVNLNLFVTKHTSFTLQVINGKAAVAAVSVLIFRNPTVMG
 IAGCVITMGVLYLEAKRSKSLK"
 complement(join(17619..17885,17959..18182,18485..18619,
 18717..18789,18871..18897,19006..19242))
 /note="similar to Cu/Zn-superoxide dismutase copper
 chaperone precursor gb|AAD52685.1; similar to ESTs
 dbj|AV442631.1, gb|BE038022.1, dbj|AV441271.1,
 dbj|AV530707.1, dbj|AV561064.1, and dbj|AV561353.1"
 /codon_start=1
 /evidence=not_experimental
 /product="t12c24..6"
 /protein_id="AAF88100.1"
 /db_xref="GI:9502393"
 /translation="MASILRSVATTSVAANAASAPIAAFSSSSSSSTMPKQSLN
 FSPLSRSSPRLIGLSRSVSSPAPATLISDNLNIOEDRAMQQLTERPVATCEGVA
 AVKNLEITIEIEVEVDLSNOVVRILSSSVKAMTQALEDTGRARILGGVQDFL
 VSAVAERKGPDIIGVNAFQVSMELARIEANFGLSGTSHWCINIEGDLTNGAAS
 GSLYNPQDQGTPELDGLGLEADKNEAFVSGKKEKLVADLIGRAVAVVYKTDNKK
 SGGLTAIVIRASAGVGENYKLCSDGDTVTEATNDFVASKY"
 complement(join(19464..19818,19915..20047,20203..20316,
 20580..20642,20957..21104))
 /note="hypothetical protein"
 /codon_start=1
 /evidence=not_experimental
 /product="t12c24..7"
 /protein_id="AAF88099.1"
 /db_xref="GI:9502392"
 /translation="MARKISKEVEYKIKDDGDFDRRLKTIIRLKENVSPUSPSS
 SHCSIKLQVIMVLVCKDFELGRLSVSLTNVMDILIAIVTFEESLYLSIEMA
 LSLAOLAVTGLQEDLRNMNIVSVKSTSLKRNQNMKTROLSDAIEFVSKMSLSOL
 SDGLWGIIRSEDKNEIRETVOSVYATLSPNCKRSTASREMHKIPPKKEARTD
 FNTSPASKROKELIKGAYENDKFGACSDVDYDNGGEANNDDEDEBELPPGG"
 join(23302..23622,23752..24150)
 /note="hypothetical protein"
 /codon_start=1
 /evidence=not_experimental
 /product="t12c24..8"
 /protein_id="AAF88076.1"
 /db_xref="GI:9502369"
 /translation="MNPDSLSLTPNFAVENDLDFSSLTIPSTAVSFQEPKPCNPVH
 SAGIENKRONCEFTMTLSEIMKDDDEPKNRAHKELEORQENSLFKILRYLP
 SOYIKGRSSADHLEAVNYIKDLQKIKVESEKDRIRKSTIHPSSGFEISILAS
 STGCVGHTNIAVVRPCLIGLELVSCNNHESCLSVLQLAQEQFNIVSCISTR
 LHQGFHTIASEVLIYNLSI"
 join(24991..25437,25649..26173)
 /note="similar to D-3-phosphoglycerate dehydrogenase
 sp|P35136|SERB_BACSV; similar to EST gb|T245682.1"
 CDS
 Query Match 3.3%; Score 128; DB 12; Length 106142;
 Best Local Similarity 50.0%; Pred. No 3.5e-11;
 Matches 402; Conservative 0; Mismatches 395; Indels 7; Gaps 3;
 Oy 292 gaagaataagaaacacgatcaactcaactcgggataaalcgggttggccattcttg 351
 111111 111 111 111 111 111 111 111 111 111 111 111 111 111
 Db 92868 GGAGAAATCTTAATGACAGAGGAACCTCCAGTTCATGGTTCCTTTACTCG 92927
 111111 111 111 111 111 111 111 111 111 111 111 111 111 111
 Oy 352 gtgaacacatcggtatcattaaacgcgtacacgcgcaacaactcgggtactcaatgaac 411
 111 111 111 111 111 111 111 111 111 111 111 111 111 111
 Db 92938 CGGAGAGTATCCAAATCTTCAAGCCAAACAACTTCAGACATCCCTCTTATCAAG 92987
 111 111 111 111 111 111 111 111 111 111 111 111 111 111
 Oy 412 aacatgctccaagtaacaacaacatctccaacaa-aaactcaaaaaataaactctcg 469
 111 111 111 111 111 111 111 111 111 111 111 111 111 111
 Db 92988 AGAGAGTTAAGAGTAAGTATCTCTCTCATTTCAAAACAAATGCTATTATCTTAT 93047
 111 111 111 111 111 111 111 111 111 111 111 111 111 111
 Oy 470 ttctgaaattgactaaatgcttattactaacggtatggttaagatatagatcgaac 529
 111 111 111 111 111 111 111 111 111 111 111 111 111 111
 Db 93048 AATGCTGCTTTTAAAGAGGATGATATGAGCAGTATGGCCCAATTTTCAAGCCAAAT 93107
 111 111 111 111 111 111 111 111 111 111 111 111 111 111
 Oy 530 ttgcttgagaagacacacatcgatcagcgtcgtgacttaatatagatcatatcaac 589
 111 111 111 111 111 111 111 111 111 111 111 111 111 111
 Db 93108 CTGCTGGGAGACCGATTATGTATCAACAGATGCTGATTGAGTTATTTGTGTAAAC 93167
 111 111 111 111 111 111 111 111 111 111 111 111 111 111

| | | | |
|------------|--|--|-------------|
| Oy | 590 | aacgaagaagacccttggatctgagttaactccttagaaglatagtgaggatctcttgaggaa | 649 |
| Db | 93168 | CAGAAGGACACTTGGTTCCATCCAAGTTTGATATCCAGACACTTTTACCACAACTCTTGGGAAG | 93227 |
| Oy | 650 | tgtgcatactcttgttcctgtgtgtgtgacalcatagatagatalagagaatctcgcttaac | 709 |
| Db | 93228 | AAGAATGGGGGTTCATTACATGATGGTTTCATGTACAAGTAACGTAAATAACATGGTTTGACT | 93287 |
| Oy | 710 | ttcttaagtcacgcacgctctcttagaactctcttacttaagaatglttgagaagacatacttg | 769 |
| Db | 93288 | CTCTTTGGCCATGATNGTCTCAAGAAGATGCTTCTCOTCAAGTAGAAAATGACCTGCCAATAG | 93347 |
| Oy | 770 | ttgtctctgattctcttggcaacaacacttatctctctgcaccaagagcccaaaag | 829 |
| Db | 93348 | AGGTTGAGAGCTTTGGTGCATAATCACAGATTCACTAGAACCTCAAAGATGCACACCGCAAGCGTG | 93407 |
| Oy | 830 | gttttatttattacotttattcttgctaacttttttgtttatgaaactttaagttct | 889 |
| Db | 93408 | AGTTTCCATCTCAGAAATPMAACATTTTATATATGTTTCAGTCTATAAATGTTTTATACTT | 93467 |
| Oy | 890 | aactttttttttttaaattgaacagttlaagtttaacttaatbgggaagcatatatag | 949 |
| Db | 93468 | GTCCTPACATGTTAGTA--TGATCACATATATATTTGATCTCACCGGGAAGAAATGATGAG | 93525 |
| Oy | 950 | tatgatcccttgagaagaagaacagacaaatataagaagaagatgtaacttcaatgaa | 1009 |
| Db | 93526 | CCATGATATCC---AGACAAAGTCATCATGACAAATCTAAGGCGMAACTTTGTTGCTTTATACA | 93582 |
| Oy | 1010 | aggagttgctctctgcctctctaaatctaacaggaactgtatcataaagctcttcagtt | 1069 |
| Db | 93583 | GGGATTTGATCTCTTCCCTTTTGATATCCACAGGACAGCTTATCCAAATATGTCACAGGT | 93642 |
| Oy | 1070 | aacttatttttttttgctgataa | 1093 |
| Db | 93643 | GAATPATCAACGTTTCATGCTTCANA | 93666 |
| RESULT | 5 | | |
| LOCUS | AC012396/c | | |
| DEFINITION | Arabidopsis thaliana chromosome 1 BAC T9L24 genomic sequence, | PLN | 08-NOV-2000 |
| ACCESSION | AC012396 | | |
| VERSION | AC012396.7 | | |
| KEYWORDS | HTG. | | |
| SOURCE | Arabidopsis thaliana | | |
| ORGANISM | Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta; | | |
| | Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II | | |
| | Brassicales; Brassicaceae; Arabidopsi. | | |
| REFERENCE | 1 (bases 1 to 90459) | | |
| AUTHORS | Liu,X., Kaul,S., Town,C.D., Benito,M., Creasy,T.H., Haas,B.J., Wu,D., Maitl,R., Rensing,C.M., Koo,H., Fujii,C.Y., Uteback,T.R., Barnstead,M.E., Bowman,C.L., White,O., Niernman,W.C. and Fraser,C.M. | | |
| JOURNAL | Arabidopsis thaliana chromosome I BAC T9L24 genomic sequence Unpublished | | |
| REFERENCE | 2 (bases 1 to 90459) | | |
| TITLE | Liu,X. and Kaul,S. | | |
| JOURNAL | Direct Submission Submitted (27-OCT-1999) The Institute for Genomic Research, 9712 Medical Center Dr, Rockville, MD 20850, USA, xlinet@igrr.org | | |
| REFERENCE | 3 (bases 1 to 90459) | | |
| AUTHORS | Town,C.D. and Kaul,S. | | |
| TITLE | Direct Submission Submitted (03-NOV-2000) The Institute for Genomic Research, 9712 Medical Center Dr, Rockville, MD 20850, USA, cdtowne@igrr.org | | |
| REFERENCE | 4 (bases 1 to 90459) | | |
| AUTHORS | Town,C.D. and Kaul,S. | | |
| TITLE | Direct Submission Submitted (08-NOV-2000) The Institute for Genomic Research, 9712 Medical Center Dr, Rockville, MD 20850, USA, cdtowne@igrr.org | | |
| JOURNAL | On Nov 8, 2000 this sequence version replaced gi:11079536. | | |
| COMMENT | | | |

FEATURES

Address all correspondence to: atet@tigr.org

BAC clone 79L24 is from Arabidopsis thaliana chromosome 1. The orientation of the sequence is from Sp6 to T7 end of the BAC clone.

Genes were identified by a combination of several methods: Gene prediction programs including Genscan+ (Chris Burge, <http://CCR-081.mit.edu/GENSCAN.html>), GeneMarkHMM (Mark Borodovsky, <http://genemark.biology.gatech.edu/genemark/>), Glimmer4 (a variant of Glimmer, see Minkele Perlea, http://www.tigr.org/softdb/glimmer_hmm/glimmerm.html, and GeneSplicer (Minkele Perlea and Steven Salzberg, <http://compbio.tigr.org>), searches of the complete sequence against a peptide database and the plant EST database at TIGR (<http://www.tigr.org/cdb/cgi.shtml>). Annotated genes are named to indicate the level of evidence for their annotation. Genes with similarity to other proteins are named after the database hits. Genes without significant peptide similarity but with EST similarity are named as unknown proteins. Genes without protein or EST similarity, that are predicted by more than two gene prediction programs over most of their length are annotated as hypothetical proteins. Genes encoding tRNAs are predicted by tRNAscan-SE (Sean Eddy, <http://genome.wustl.edu/eddy/tRNAscan-SE/>) Simple repeats are identified by RepeatMasker (Arrian Smit, <http://ftp.genome.washington.edu/RM/RepeatMasker.html>).

Location/Qualifiers

1. 90459

```
repeat_region
repeat_region
```

1. .90459.
/organism="Arabidopsis thaliana"
/cultivar="Columbia"
/db_xref="taxon:3702"
/chromosome="1"
/map="ngali1"
/clone="T9L24"
complement(join(<1. .73,432. .647,958. .1023,1079. .1375,
1456. .>1980).
/gene="T9L24.27"
complement(1. .1980)
/gene="T9L24.27"
/note="similar to hypothetical protein GI:8671775 from
(Arabidopsis thaliana)"
complement(join(<1. .73,432. .647,958. .1023,1079. .1375,
1456. .1980).
/gene="T9L24.27"
/codon_start=1
/product="hypothetical protein, 3' partial"
/protein_id="AAG30966.1"
/db_xref="GI:11120786"
/translation="MRYSGQSIITAIISLFTLSITYIHRVQARPEVLVSNEDLNGGFNDN
GAYSEDFDEGESEEEELDPGSMRIPEFNESTVHASAPQYISGLKILIASA
EGMTTAAEIVSEIDSSAGDHAQSVGIVGICMMETSRSSILHHNFHAAKNDK
MOSKMTLAAEFTIRNDKRAVGLYLAELAEVANSFLIDSSMPAPVLIHIGTEENK
ALRSRGEDEEDPDITLEYQAEKNSVAMKRIGLFTYTGIRGIRRHAAALVWFSKAE
KGLFGLVYRGVQDKRYNTAREYFEAANNEDPSGHIWGLVLYLKGIVKDKVRRHAK
KFFFLAANAAGQAFVOLAAMFTGVLTKNLEMAITTFYKIVAEKGPWSSLSRMALAEK

#<2516. .>2854
/gene="T9L24.28"
2516. .2854
/gene="T9L24.28"
/note="contains similarity to lipid transfer protein
GI:2627141 from (Picea abies)"
2516. .2854
/gene="T9L24.28"
/codon_start=1
/product="hypothetical protein"
/protein_id="AAG30968.1"
/db_xref="GI:11120788"
/translation="MEINRKLAVVAVVLYVSEATAGCNQLTACLQKLLPCOPYV
HSLNPPSPSCGGMKRIVKADPAICIAFNNPEVLKALNLTRENNALLPACGVPNDPC
VSIKSIASE"

#complement(3212. .3233)
/protein_id="AT_112111"
complement(3974. .4011)

```

mRNA      /rpt_family="AT-rich"  
          join(<4436..4769,4878..>4948)  
gene      /gene="T9L24.29"  
          4436..4948  
          /gene="T9L24.29"  
          /note="contains similarity to lipid transfer protein  
GI:2627141 from (Picea abies)"  
join(<4436..4769,4878..4948)  
CDS       /gene="T9L24.29"  
          /codon_start=1  
          /product="hypothetical protein"  
          /protein_id="AAG30973.1"  
          /db_xref="GI:11120793"  
repeat_region repeat__region  
          /translation="MNITRLGVVTTVIILYSVYTAQFGSDVOAMCAKIMPCOP  
YLHSLPPPLCCNMFKQIAEKDVSCLTARFHPDLRLFLALTKENAIKIILDSCGINH  
DPSVNCKNTNPPSPALPEPATSGMHLPYPYN"  
          complement(5193..5216)  
          /rpt_family="AT-rich"  
          /rpt__family="AT-rich"  
          complement(5484..5515)  
mRNA      /rpt_family="(C)n"  
          join(6312..6594,6690..7310)  
gene      /gene="T9L24.30"  
          6312..7310  
          /gene="T9L24.30"  
          /note="contains similarity to diphosphoinositol  
polyphosphate phosphohydrolase GI:3978224 from (Homo  
sapiens)"  
join(6464..6594,6690..7092)  
CDS       /gene="T9L24.30"  
          /codon_start=1  
          /product="unknown protein"  
          /protein_id="AAG30977.1"  
          /db_xref="GI:11120797"  
          /translation="MNNNPISKTIIVPTPSEKVLSVSRGRDLORYNTAGRVONG  
CEPVYKHGGELTEVLLISAKKGGMGLPKGMEIDSEFEALRETFEAGTYGO  
LEESYGKKQYKSKRTMTIHDSMPPLVSSQGFETWPSEEFQRKWSLSEIIECONS  
WNRKELEAFINRKCCTO"  
join(<7586..7832,8692..8760,8845..8951,9049..>9171)  
mRNA      /gene="T9L24.48"  
          7586..9171  
          /gene="T9L24.48"  
          /note="similar to RNA binding glycine rich protein (RCP  
1a) GI:36789 from (Nicotiana sylvestris)"  
join(<7586..7832,8692..8760,8845..8951,9049..9171)  
CDS       /gene="T9L24.48"  
          /codon_start=1  
          /product="RNA-binding glycine-rich protein, putative"  
          /protein_id="AAG30979.1"  
          /db_xref="GI:11120799"  
          /translation="MAISLGIRVVVPSCGISGDRFLIPNSAICSVSCGRINGTVIS  
ARRRDIGVLISCLTSDDSSPSSSGSKTLVYGSFETEDTLDTDFEPQFN  
LIHMVMVDKVAANRPKGFAPLRFTBEEMAKAIGCMGKFLDGRVITEAKIRSDMS  
RAKPRDPKQSKRPRTFRW"  
          <9854..>10321  
          /gene="T9L24.31"  
          9854..10321  
          /gene="T9L24.31"  
          /note="predicted by genemark hmms"  
          9854..10321  
CDS       /gene="T9L24.31"  
          /codon_start=1  
          /product="hypothetical protein"  
          /protein_id="AAG30982.1"  
          /db_xref="GI:11120802"  
          /translation="MAIKRKQINESSTIGFNLSIKRRHNKTTHDKPFLVYSULINSKA  
SOPSTDEIIIDDDEKENVILEPKRKPRUGSMWDNVAEVDELCTCVKSGPPAKEDN  
GVASADVVSFSTDLHEMKKNKINEDKRS.LPEKDDESTVGESNTKNGMMN"  
          complement(11389..11410)  
          /rpt_family="AT-rich"  
          /rpt__family="AT-rich"  
          complement(11574..11604)  
          /rpt_family="AT-rich"  
          12616..13881  
          /gene="T9L24.32"  


```

[illegible]

| | | | |
|------------|---|--|---------------------|
| OY | 1010 | aggactcttcgtccgcctaaccataaccagaactgcatcattacaaagcttcaagt | 1069 |
| | | | |
| Db | 79025 | TGGAGCTGCTGTCTGTTCCATTGACTTCCTCGTTCACATACACAAGCAATGAAGCT | 78966 |
| OY | 1070 | acaattt 1075 | |
| | | | |
| Db | 78965 | AAGTTT 78960 | |
| RESULT | 6 | | |
| AB005237/c | | | |
| LOCUS | AB005237 | 87835 bp | DNA PLN 27-DEC-2000 |
| DEFINITION | Arabidopsis thaliana genomic DNA, chromosome 5, Pl clone:MJ3. | | |
| ACCESSION | AB005237 | BA000015 | |
| VERSION | AB005237.1 | GI:2264309 | |
| KEYWORDS | | | |
| SOURCE | | | |
| ORGANISM | Arabidopsis thaliana (strain:Columbia) DNA, clone.lib:Mitsui P1 clone:MJ3. | | |
| REFERENCE | Arabidopsis Thaliana Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsi | | |
| AUTHORS | 1 (siles) | | |
| TITLE | Sato,S., Kotani,H., Nakamura,Y., Kaneko,T., Asamizu,E., Fukami,M., Miyajima,N. and Tabata,S. | | |
| JOURNAL | Structural analysis of Arabidopsis thaliana chromosome 5. I. Sequence features of the 1.6 Mb regions covered by twenty physically assigned P1 clones | | |
| MEDLINE | DNA Res. 4 (3), 215-230 (1997) | | |
| REFERENCE | 97471969 | | |
| AUTHORS | 2 (bases 1 to 87835) | | |
| JOURNAL | Nakamura,Y. | | |
| TITLE | Direct Submission | | |
| COMMENT | Submitted (02-JUN-1997) Yasukazu Nakamura, Kazusa DNA Research Institute, Department of Plant Gene Research, 1532-3, Yana, Kisarazu, Chiba 292-0812, Japan (E-mail:yinakamu@kazusa.or.jp, Tel:81-438-52-3935, Fax:81-438-52-3934) | | |
| | Address for correspondence: kaos@kazusa.or.jp | | |
| | http://www.kazusa.or.jp/kaos/cgi-bin/agd_graph.cgi?re=MJ3 | | |
| | For the latest information on annotation of this clone, please see | | |
| | Genes with similarity to proteins in the databases are described in | | |
| | 'product' or 'note' qualifiers. Genes that have no significant | | |
| | protein similarity are described as 'unknown protein'. | | |
| | The software programs used to predict genes include: Grail | | |
| | (Informatics Group, Oak Ridge National Laboratory, | | |
| | http://compbio.ornl.gov/Grail-1.3/), | | |
| | GENSCAN (Chris Burge, MIT, http://CCR-081.mit.edu/GENSCAN.html), | | |
| | NetGene2 (S.M. Hebsgaard, et al., CBS, Technical University of | | |
| | Denmark, http://www.cbs.dtu.dk/services/NetGene2/) and | | |
| | SplicePredictor (Volker Brendel, Stanford University, | | |
| | http://gremlin1.zoool.iastate.edu/cgi-bin/sp.cgi). | | |
| | Genes encoding tRNAs are predicted by tRNAscan-SF | | |
| | (Sean Eddy, Washington University School of Medicine, St. Louis, | | |
| | http://genome.wustl.edu/eddy/tRNAscan-SF). | | |
| | This sequence may not be the entire insert of this clone. It may be | | |
| | shorter because we remove overlaps between neighboring submissions. | | |
| | The 5' clone is MOP10 and the 3' clone is K18U17. | | |
| FEATURES | | | |
| Source | | | |
| | 1..87835 | | |
| | /organism="Arabidopsis thaliana" | | |
| | /strain="Columbia" | | |
| | /db_xref="taxon:3702" | | |
| | /chromosome="5" | | |
| | /clone="MJ3" | | |
| | /clone_lib="Mitsui P1" | | |
| | join(63)..191,310..622,711..1111,1255..1308,1383..1487, | | |
| | 1590..1664,1757..1801,1946..2074,2184..2288,2374..2442, | | |
| | /note="gene_id:MJ3.1 | | |
| | sp P38558" | | |
| | /codon_start=1 | | |
| | /evidence=not_experimental | | |
| | /product="tubulin gamma-2 chain" | | |

1

```
/protein_id="BAB09656.1"
/db_xref="gi:9759087"
/translation="MPEKITIYLVGGCGGNGQMGFPWKQCLNENHISKDGLIEDPATOG
GDRKDVPEYQADDDHYIPRALILDEPVIINGIONGERNLYNHNELPLSDHGGAAGN
NMASGYHOGKACVEEIDMDIDREADSDSLGFEVLCHSIAGTSGMGSYLLETFLNDR
YSKTIQVQSYSEPNQMETSDVYVQPYNSLTLNRLTLNADVVLDNTALMLRIVERL
HLNTEPFOETSLVSTVKSASTTTLRYPGYNNDLGLGLSLPTPCGHFLMTGTPEL
TVERQANIRIKETVLADWRRLLOTKNIMVSSYANKNESSOKYIILNITIOGVDPPRO
VLESIQRIAREKRLVNEFDMGSPASIOVALSKSPYVQVISHRSGTGLMASHSTRLEP
CLSDQILRRKQOAFLDNTRKPFMEADNDLSEFDESRLIESLVDEYACESPDIYKMG
MEDPQOLTGGCMNSGVADPKLAF"
3264..4736
/notes="contains similarity to cationic amino acid
transporter 1
gene_id:MJ33.2"
/codon_start=1
/evidence="not_experimental"
/protein_id="BAB09657.1"
/db_xref="gi:9759088"
/translation="MTESSPNLDSASOKPRISTENPPPHHSISIGVTGCDPATSPR
RTVNOIKITIVLPVFLIFEVSGGPEIEDSVKACGLVILAGFIVPPLTWSIPEDAL
ITAEINGENPENGIVVWVTLAMGPWGEFOGWKVMQSLDNLAPLFLDYLRKSDI
PILSGIIPRAVAIIVLIFALTYLNRGSLISYGAVALIGVSLIPFYVMGMSYIPKRL
PSRMLVSRGKKGVMKWSLYLNTLPMNLNMPQSYLDEVEVNPSTKIPTRALFYVLLY
VESYFPEPLTNGEATLIDOKLMTGTFADIDKIVGGVALGQVILISMAATSMGFLAB
MSSDFOLTLGMAERGMLEPYVAKRSRYTPVAGLIFPSGVIILISSTFOGIVAAENL
LYCEWVLEETFEVRLRRKRYPAASRPRIIPVGVLGSYLMCIPTPTVLIGVIMAFNTLKV
ALVSLAAVILVLOPCLCKOVERKGMLEKFTSSHLPNLMF"
5329..6300
/notes="gene_id:MJ33.3
similar to unknown protein
sp|P03515"
/codon_start=1
/evidence="not_experimental"
/protein_id="BAB09658.1"
/db_xref="gi:9759089"
/translation="MEEALSEALSDFDEDELVASLADAIAATSPKPLVERSLIAHMA
ELFKRTKTLPRAEVSRTEFFRAVSTIGLNGNTRKDOLEPTCAGAKIIRKVAITGIE
RDKGQRKVLVINKTVYTESRMALIFPPHMASQLMVSKESELVNRPASFPNDANFPAQTHR
LPIALHSGGFGGLIPRESDDMEFIDLICMCHSAFAEPFSLITPOEORSFALVLENOK
RYTEAARSGCNSDENRIYVILRFGLISDSLDMKLKVAAVSTLSAQEBAVSYTRDS
FTYVEDRKIYNDDAKKMDLELLKKAKRIYDASSMKSQVAAVDSLQGV"
complement(7255..7632)
/notes="gene_id:MJ33.4
similar to unknown protein"
/codon_start=1
/evidence="not_experimental"
/protein_id="BAB09659.1"
/db_xref="gi:9759090"
/translation="MDQLSEKOIWIYHIIPEYQTEKSFYACVRRSGMKIPDTFFPPHN
LSARPTORSIKFGDWSGSCLYDPDIYDGKISIQVGCVLPPEAEIIGCVCLILF
LIKILSLRYKMSHSHSPVSCDL"
9650..10282
/notes="gene_id:MJ33.5
similar to unknown protein"
/codon_start=1
/pseudo
/evidence="not_experimental"
complement(goln(11836..11997,12102..12191,12278..12653,
12819..13009,13236..13403,13534..13696,14495..14634,
15024..16196))
/notes="contains similarity to unknown protein
gb|AAF47871.1
gene_id:MJ33.6"
/codon_start=1
/evidence="not_experimental"
/protein_id="BAB09660.1"
/db_xref="gi:9759091"
/translation="MTNMGATATTEFRWKKSPPOPSOROPISDSDSGSGSNHQRH
HNDLSNTEFAYLDCSSSSPSIDLATIQSFLLASSSGAGVCLICLERIKRTPPTS
CTSSCFEAFHFLCTIOSMARQCIDQAAAVTRPSSNPTPEBAVNMCPKCRSSYOKSKIT
```

[illegible]

| Query Match | 2.4% | Score | 92.8 | DB 9: | Length | 4937; |
|-----------------------|---|--------------|-------------|-------------|--------|--------|
| Best Local Similarity | 53.1% | Pred. | NO.1.6e-05; | | | |
| Matches | 225; | Conservative | 0; | Mismatches | 192; | Indels |
| | | | | | | 7; |
| | | | | | | Gaps |
| | | | | | | 1; |
| Db 27234 | CCGGTACGGTTTCGGTTTTCATACGACATCTTTTGGTGAACGACGAGATTTCCTACGCTGA | 27175 | | | | |
| QY 562 | tgctggacttaataagatcattcaataatacaaaagaaaggaagcctcttgaaatgtagtacc | 621 | | | | |
| Db 27174 | CCCGGAACGAACCGGTTTGTTCTTCAGAACGAAGGAGCTTTTGGAGTCTTCATTC | 27115 | | | | |
| QY 622 | tagaagatcagaatggagatctctgggaaatgagcagatgctgcttgcttgatgagatgca | 681 | | | | |
| Db 27114 | TGCTTCATTTTGAACCTTTTGGGGAACACTCTCTGCTCTTATGAAAGGTTCTTTGCA | 27055 | | | | |
| QY 682 | tagagataagaaagatctcgcttaactctctaagtcacagcctctagaaactatct | 741 | | | | |
| Db 27054 | TAAAGGTATGCACTCTCTACCATGAGCTTTGCTATTTCTTAATCAATTAAGACCATCT | 26995 | | | | |
| QY 742 | acttaagaatgltgagagacatacttgcttgcttgatctcttgagcaaaactctat | 801 | | | | |
| Db 26994 | CATGCTTGATATATGACCGGTTAGTCCGGTTTATCTGATCTTGCG-----TCTTTC | 26942 | | | | |
| QY 802 | ttctctgctcaagcagcgagcgaagaagcttttaacttttaactcttaatttgcaatc | 861 | | | | |
| Db 26941 | GGTTCTCTCTCTGGAAGAGCAAGCAAAAGATACCAAAAAATTCCTTATTCATAAAC | 26882 | | | | |
| QY 862 | ttttgttataagatccttaagatcttaactcttttttttaacttgaaacagttacgt | 921 | | | | |
| Db 26881 | ATTATATATATTTATTTATTTGCGGCTTCTTCACCTATGTTTTTTTATATATAAATAAT | 26822 | | | | |
| QY 922 | ttaatc 927 | | | | | |
| Db 26821 | AAAAAC 26816 | | | | | |
| RESULT 7 | | | | | | |
| LOCUS AR074364 | 4937 bp | DNA | PAT | 28-AUG-2000 | | |
| DEFINITION | Sequence 3 from patent US 5952545. | | | | | |
| ACCESSION | AR074364 | | | | | |
| VERSION | AR074364.1 | GI:10001119 | | | | |
| KEYWORDS | Unknown. | | | | | |
| SOURCE | Unknown. | | | | | |
| ORGANISM | Unknown. | | | | | |
| REFERENCE | 1 (bases 1 to 4937) | | | | | |
| AUTHORS | Koncz, C., Mathur, J., Szekeres, M. and Altmann, T. | | | | | |
| TITLE | Nucleic acid molecules encoding cytochrome P450-type proteins | | | | | |
| JOURNAL | Involved in the brassinosteroid synthesis in plants | | | | | |
| FEATURES | Patent: US 5952545-A 3 14-SEP-1999; | | | | | |
| source | location/Qualifiers | | | | | |
| | 1..4937 | | | | | |
| | /organism="unknown" | | | | | |
| BASE COUNT | 1526 a 864 c 870 g 1677 t | | | | | |
| ORIGIN | | | | | | |

| QY | 742 | actaagaagttgagacacactttgttctgtcatcttgcgaacaacatctcat | 801 |
|------------|---|--|-------------|
| Db | 1402 | CATGCTTGATATATATACCGGTTAGTCGCCGTTTAAATTTCTTGG-----TCTTCTC | 1454 |
| QY | 802 | ttctctgcctcaagcgcggccaaagaagtttatttttactcttattttgtccaatt | 861 |
| Db | 1455 | GTGTTCTCCTCATGSAACAAGCAGCAAAAAGTACCAAAAATTTCTTGCTTATCAAAAC | 1514 |
| QY | 862 | ttttgtttatgaactttagagtttagtttctaactttttttttttaaattgaacagttacgt | 921 |
| Db | 1515 | ATTATATATATATATTTATTTATTCGGCTTCTCACCCTATGTGTTTTTTTATATATAAATAAAT | 1574 |
| QY | 922 | ttaa | 925 |
| Db | 1575 | AAAA | 1578 |
| RESULT | 8 | | |
| LOCUS | ATCYP450D | | |
| DEFINITION | ATCYP450D | 4937 bp | DNA |
| ACCESSION | A.thaliana | gene cytochrome P450. | PLN |
| VERSION | X87368.1 | GI:871987 | 07-OCT-1996 |
| KEYWORDS | Cytochrome P450. | | |
| SOURCE | thale cress. | | |
| ORGANISM | Arabidopsis thaliana | | |
| REFERENCE | Eukaryota: Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta | | |
| AUTHORS | Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II, | | |
| TITLE | Brassicales; Brassicaceae; Arabidopsis. | | |
| JOURNAL | 1 (bases 1 to 4937) | | |
| MEDLINE | Szekeres,M., Nemeth,K., Koncz-Kalman,Z., Mathur,J., Kauschmann,A., | | |
| REFERENCE | Altmann,T., Redei,G.P., Nagy,F., Schell,J. and Koncz,C. | | |
| AUTHORS | Brassinosteroids rescue the deficiency of CYP90, a cytochrome P450 | | |
| TITLE | controlling cell elongation and de-etiolation in Arabidopsis | | |
| JOURNAL | Cell 85 (2), 171-182 (1996) | | |
| REFERENCE | 96200769 | | |
| AUTHORS | 2 (bases 1 to 4937) | | |
| TITLE | Szekeres,M. | | |
| JOURNAL | Direct Submission | | |
| REFERENCE | Submitted (22-MAY-1995) M. Szekeres, Inst. of Plant Biology, | | |
| AUTHORS | Biological Research Center, Hungarian Academy of Sciences, PO Box | | |
| TITLE | 521, 6701 Szeged, HUNGARY | | |
| JOURNAL | 3 (bases 1 to 4937) | | |
| REFERENCE | Szekeres,M. | | |
| AUTHORS | Direct Submission | | |
| TITLE | Submitted (22-JUN-1995) M. Szekeres, Inst. of Plant Biology, | | |
| JOURNAL | Biological Research Center, Hungarian Academy of Sciences, PO Box | | |
| REFERENCE | 521, 6701 Szeged, HUNGARY | | |
| AUTHORS | 4 (bases 1 to 4937) | | |
| TITLE | Szekeres,M. | | |
| JOURNAL | Direct Submission | | |
| REFERENCE | Submitted (26-JUN-1995) M. Szekeres, Inst. of Plant Biology, | | |
| AUTHORS | Biological Research Center, Hungarian Academy of Sciences, PO Box | | |
| TITLE | 521, 6701 Szeged, HUNGARY | | |
| JOURNAL | On Jun 28, 1995 this sequence version replaced gi:853720. | | |
| COMMENT | Location/Qualifiers | | |
| FEATURES | 1..4937 | | |
| SOURCE | /organism="Arabidopsis thaliana" | | |
| | /cultivar="Columbia" | | |
| | /db_xref="taxon:3702" | | |
| | /chromosome="5" | | |
| | /clone_lib="lambda EMBL4" | | |
| | /clone="G206" | | |
| | /map="14.3" | | |
| | 755 802 | | |
| | join(792..1483,1680..1829,1917..2165,3903..3989, | | |
| | 4084..4162,4248..4354,4446..4576,4674..4916) | | |
| | 792..1483 | | |
| | /number=1 | | |
| | join(968..1483,1680..1829,1917..2165,3903..3989, | | |
| | 4084..4162,4248..4354,4446..4576,4674..4773) | | |
| | /gene="CYP" | | |
| | join(968..1483,1680..1829,1917..2165,3903..3989, | | |
| CDS | | | |

| Query Match | Best Local Similarity | Score | DB | Length |
|--|---|-------|-----|--------|
| Matches 225; Conservative 0; Mismatches 192; Indels 7; Gaps 1; | 53.18; | 92.8; | 13; | 4937; |
| Db 502 | cagatgtagtataatagatagatcgaactgttttggagagacaacgacgtatcagctga | 561 | | |
| Db 1152 | CCGGTACGGTTCGGTTCATGACGACATCTTTTGGTATACCGACGATTTTCTCAGCTGA | 1221 | | |
| Qy 562 | tcgtgacttaatagatcatcattacaaacgaagaaagctcttgaatytagtlaacc | 621 | | |
| Db 1222 | CCCGGAACGAACCGCTTGTCTTCAAGAACGAAGCGATTTTGTGAGTGTCTTATCC | 1281 | | |
| Qy 622 | tagaagtataagtgagatctcttgggaaatggtcgatgctgttcttctgttgatgacatga | 681 | | |
| Db 1282 | TGCTTCATTGTGTAACTTTTGGGGAACACTCTCTGCTTATGAAGAAGTCTTTTCA | 1341 | | |

| | | |
|--------|--|--|
| exon | /gene="AT4g36190" /number=10 7653. .7736 | /gene="AT4g36190" /number=11 7737. .7853 |
| intron | /gene="AT4g36190" /number=11 7854. .7894 | /gene="AT4g36190" /number=12 7895. .8795 |
| exon | /gene="AT4g36190" /number=12 8796. .8921 | /gene="AT4g36190" /number=13 8922. .9011 |
| intron | /gene="AT4g36190" /number=13 9012. .9185 | /gene="AT4g36190" /number=14 9186. .9264 |
| exon | /gene="AT4g36190" /number=14 9265. .9450 | /gene="AT4g36190" /number=15 9451. .9551 |
| intron | /gene="AT4g36190" /number=15 9552. .9668 | /gene="AT4g36190" /number=16 9669. .9767 |
| exon | /gene="AT4g36190" /number=16 9768. .9821 | /gene="AT4g36190" /number=17 9822. .9900 |
| intron | /gene="AT4g36190" /number=17 9901. .9972 | /gene="AT4g36190" /number=18 9973. .10582 |
| exon | /gene="AT4g36190" /number=18 10583. .10657 | /gene="AT4g36190" /number=19 10658. .10726 |
| intron | /gene="AT4g36190" /number=19 10727. .10865 | /gene="AT4g36190" /number=20 10866. .10979 |
| exon | /gene="AT4g36190" /number=20 10980. .11095 | /gene="AT4g36190" /number=21 11360. .11432 |
| intron | /gene="AT4g36190" /number=21 11536. .11982 | /gene="AT4g36200" /number=1 11982. .11982 |
| exon | /gene="AT4g36200" /number=1 11982. .11982 | /gene="AT4g36200" /number=1 11982. .11982 |
| gene | complement(11536. .11982) | complement(11536. .11982) |

| Query Match | 1.9% | Score 73.6 | DB 13 | Length 198750 |
|------------------------|--|--------------------------|-----------|---------------|
| Blast Local Similarity | 47.5% | Prod. No. 0.013 | | |
| Matches 374 | Conservative 0 | Mismatches 384 | Indels 30 | Gaps 4 |
| | /gene="AT4g36200" | complement(11536..11982) | | |
| | /gene="AT4g36200" | | | |
| | /codon_start=1 | | | |
| QY 491 | tgattattctcaagtgatgagatataatagatgaactgttttggaagcaacagatc | 550 | | |
| Db 71942 | TGTTAAACAATATATAGTACGGAAAGTGTTCAAACGAAACATATATAGGACACACATCATTA | 71883 | | |
| QY 551 | gtatcagctgtagctgagacttaatagatcaltacataaacaaggaagagctcttgaa | 610 | | |
| Db 71882 | ATATCAACCGATGACAGAGAGTAAATAAAGTGTGCTCCCAAAACATGAGGACACATTTGTC | 71823 | | |
| QY 611 | tgtatgtagtctagaagatagtggtgagttcttllggaatggttgatgctgttctgtt | 670 | | |
| Db 71822 | CCTGCATACCTCTAAATCAATTAACGGAACCTACTTTGGAGAAACCTCATTTCTACATCAAT | 71763 | | |
| QY 671 | ggtgacatgcatagagatgagaagatctccgcgttaactcttaagtcacagcaagcttc | 730 | | |
| Db 71762 | GGACCTCATCAAAAAAGGCTTCACACGCTCATTTGGCGGCTTCCTACGATCTCCTACCTC | 71103 | | |
| QY 731 | agaactattctactaaagatgltgagagacatacttgttctgtcttgatcttgagcaa | 790 | | |
| Db 71702 | AAAGACCGATACACGACGACATTGAGACGCTCGGTTGTTTCACATTTTGGCGCTTGAGCT | 71643 | | |
| QY 791 | caaaactctattctctcgtgccaagaagagccaaagaagtttatttattctttat | 850 | | |
| Db 71642 | CACCTTCATCTGTTGCATAGTTTCAGAGTATGAGATCAAAAAAGTATGCTCTTAGAATAATAT | 71583 | | |
| QY 851 | t-----ttgttaaatctttttgttttgaactctttagagttctta | 890 | | |
| Db 71582 | TAGTAGCAAACTCACACGTACGTACTAGCTACATGCTATATACACATTTTAAATTGGTA | 71523 | | |
| QY 891 | actttttttttttaaattgaaacagtttaacgtttaaactaaatgagcgaagacataatgagt | 950 | | |
| Db 71522 | GAATGTTTGTGTTGTTGTTGTGTGATGATGACGTTTGAGATATTAGTAAAGTGTGATGAGC | 71463 | | |
| QY 951 | atggaactctgagaagaagaacaagacaatlaaagaagaagtagtaacttctcagaa | 1010 | | |
| Db 71462 | ACATCTCCTGG--TGAAGATATGAACATTTCTCAAACTTGAGTTGGAAGAAATTCATCAAA | 71406 | | |
| QY 1011 | ggaagtgtcctcgtccctctaaactaccagaagactgtcttaccataagctctcaggtta | 1070 | | |
| Db 71405 | GATTGATTGTTGATCCCAATCAAAATTCCTTGACCTTAAGACTCTACAAATTCCTTAAAGGTA | 71346 | | |
| QY 1071 | catattatttttttgcgtgaagtcacaaactcactatlaagtttttatttatttt | 1130 | | |
| Db 71345 | AA-----ATATTTCCTTAAATATACATGACGATATCAATATCTGATTTTCAATTTGTTAGT | 71292 | | |
| QY 1131 | atgtgttaataataaatal-ctaaatggttlytltgagtcagagcaagaaatltgaagttc | 1189 | | |
| Db 71291 | GTGTAAATACATGAGGTTTAAATTAATCTCAATTCATGCGAGGAAAGAGAGGTTTAAATGAATG | 71232 | | |
| QY 1190 | attgagagaataatgagaagagaanaatttgatatacaaggagaagaagatcagaagaagaa | 1249 | | |
| Db 71231 | GTAATAAAGGTTGTGAGAGAGACAAATGGCGATGACACAGACGCTTCCTGGCAATATGAC | 71172 | | |
| QY 1250 | gaagtga 1257 | | | |
| Db 71171 | GTGTCGCA 71164 | | | |
| RESULT 15 | | | | |
| PFSC04088 | standard; DNA; INV; 21202 BP. | | | |
| XX AL010273; | | | | |
| XX AL010273.1 | | | | |

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 17, 2001, 19:18:28 : Search time 431.79 Seconds
(without alignments)
5655.317 Million cell updates/sec

Title: US-09-502-426-1_COPY_3000_6888

Perfect score: 3889

Sequence: 1 atgaagctcattgtagtt.....cctagttcagcccaagctt 3889

Scoring table: IDENTITY_MNC

Gapop 10.0 , Gapext 1.0

Searched: 730101 seqs, 313950809 residues

Total number of hits satisfying chosen parameters: 1460202

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database :
1: N.Geneseq_0601.*
2: /SIDS1/gcgdata/geneseq/geneseqn/NA1980.DAT.*
3: /SIDS1/gcgdata/geneseq/geneseqn/NA1981.DAT.*
4: /SIDS1/gcgdata/geneseq/geneseqn/NA1982.DAT.*
5: /SIDS1/gcgdata/geneseq/geneseqn/NA1983.DAT.*
6: /SIDS1/gcgdata/geneseq/geneseqn/NA1984.DAT.*
7: /SIDS1/gcgdata/geneseq/geneseqn/NA1985.DAT.*
8: /SIDS1/gcgdata/geneseq/geneseqn/NA1986.DAT.*
9: /SIDS1/gcgdata/geneseq/geneseqn/NA1987.DAT.*
10: /SIDS1/gcgdata/geneseq/geneseqn/NA1988.DAT.*
11: /SIDS1/gcgdata/geneseq/geneseqn/NA1989.DAT.*
12: /SIDS1/gcgdata/geneseq/geneseqn/NA1990.DAT.*
13: /SIDS1/gcgdata/geneseq/geneseqn/NA1991.DAT.*
14: /SIDS1/gcgdata/geneseq/geneseqn/NA1992.DAT.*
15: /SIDS1/gcgdata/geneseq/geneseqn/NA1993.DAT.*
16: /SIDS1/gcgdata/geneseq/geneseqn/NA1994.DAT.*
17: /SIDS1/gcgdata/geneseq/geneseqn/NA1995.DAT.*
18: /SIDS1/gcgdata/geneseq/geneseqn/NA1996.DAT.*
19: /SIDS1/gcgdata/geneseq/geneseqn/NA1997.DAT.*
20: /SIDS1/gcgdata/geneseq/geneseqn/NA1998.DAT.*
21: /SIDS1/gcgdata/geneseq/geneseqn/NA1999.DAT.*
22: /SIDS1/gcgdata/geneseq/geneseqn/NA2000.DAT.*
23: /SIDS1/gcgdata/geneseq/geneseqn/NA2001.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description |
|------------|-------|-------------|--------|-------|-------------|
| 1 | 3889 | 100.0 | 6888 | 21 | AA59599 |
| 2 | 200 | 5.1 | 936 | 22 | AAF58252 |
| 3 | 200 | 5.1 | 936 | 22 | AAF58254 |
| 4 | 200 | 5.1 | 936 | 22 | AAF58257 |
| 5 | 200 | 5.1 | 936 | 22 | AAF58259 |
| 6 | 200 | 5.1 | 936 | 22 | AAF58262 |
| 7 | 200 | 5.1 | 936 | 22 | AAF58255 |
| 8 | 198.8 | 5.1 | 936 | 22 | AAF58252 |
| 9 | 198.8 | 5.1 | 936 | 22 | AAF58254 |
| 10 | 198.8 | 5.1 | 936 | 22 | AAF58257 |
| 11 | 198.8 | 5.1 | 936 | 22 | AAF58259 |

| | | | | | |
|----|-------|-----|-------|----|----------|
| 12 | 198.8 | 5.1 | 936 | 22 | AAF58262 |
| 13 | 198.8 | 5.1 | 938 | 22 | AAF58255 |
| 14 | 92.8 | 2.4 | 4937 | 18 | AAT85307 |
| 15 | 87.6 | 2.3 | 1608 | 18 | AAT85306 |
| 16 | 87.6 | 2.3 | 1646 | 21 | AAC48157 |
| 17 | 87.6 | 2.3 | 1649 | 21 | AAC47986 |
| 18 | 71.4 | 1.8 | 244 | 22 | AAF58238 |
| 19 | 66.6 | 1.7 | 1398 | 21 | AAC48714 |
| 20 | 66.2 | 1.7 | 244 | 22 | AAF58238 |
| 21 | 57.4 | 1.5 | 605 | 17 | AAT31530 |
| 22 | 55.4 | 1.4 | 940 | 21 | AAC37912 |
| 23 | 55.4 | 1.4 | 1673 | 21 | AAC39858 |
| 24 | 54.2 | 1.4 | 665 | 21 | AAZ31956 |
| 25 | 52.8 | 1.4 | 49999 | 20 | AAZ31902 |
| 26 | 52.6 | 1.4 | 2601 | 21 | AAC48719 |
| 27 | 52.4 | 1.3 | 2300 | 20 | AAZ00686 |
| 28 | 52.2 | 1.3 | 1667 | 21 | AAC35582 |
| 29 | 52.2 | 1.3 | 26811 | 20 | AAZ20253 |
| 30 | 51.8 | 1.3 | 1697 | 21 | AAC77781 |
| 31 | 51.6 | 1.3 | 1251 | 19 | AAV34685 |
| 32 | 51.4 | 1.3 | 1493 | 21 | AAZ43802 |
| 33 | 51.4 | 1.3 | 1733 | 21 | AAZ37036 |
| 34 | 51.4 | 1.3 | 1733 | 22 | AAF54238 |
| 35 | 51.4 | 1.3 | 1734 | 22 | AAF92083 |
| 36 | 51.4 | 1.3 | 1734 | 22 | AAC87038 |
| 37 | 51.2 | 1.3 | 2846 | 21 | AAZ65034 |
| 38 | 51.2 | 1.3 | 2846 | 22 | AAF92076 |
| 39 | 51.2 | 1.3 | 2846 | 22 | AAF44180 |
| 40 | 51 | 1.3 | 6644 | 20 | AAZ33181 |
| 41 | 51 | 1.3 | 7372 | 20 | AAZ33182 |
| 42 | 51 | 1.3 | 7797 | 20 | AAZ33180 |
| 43 | 51 | 1.3 | 7996 | 20 | AAZ33184 |
| 44 | 50.2 | 1.3 | 19124 | 18 | AAT72682 |
| 45 | 50.2 | 1.3 | 19124 | 21 | AAZ96287 |

ALIGNMENTS

| | | |
|----------|---|-------------------------|
| RESULT 1 | AA59599 | standard; DNA; 6888 BP. |
| ID | AA59599 | standard; DNA; 6888 BP. |
| XX | AA59599; | |
| AC | AA59599; | |
| XX | 14-NOV-2000 | (first entry) |
| DE | DNA encoding a cytochrome P450 enzyme designated DMF4. | |
| XX | DMF4: cytochrome P450 enzyme; brassinosteroid; 22alpha-hydroxylation; | |
| KW | plant phenotype; cell elongation; ss. | |
| XX | Arabidopsis sp. | |
| OS | Arabidopsis sp. | |
| XX | key | |
| XX | promoter | |
| FT | TATA signal | |
| FT | CDS | |
| FT | exon | |
| FT | intron | |
| FT | exon | |
| FT | intron | |
| FT | exon | |

| | | |
|----------------------|------------|---------------------|
| Oligonucleotide D2 | 1.3203 | Location/Qualifiers |
| Oligonucleotide D1 | 3060..3125 | /*tag= a |
| Arabidopsis thaliana | 3203..6110 | /*tag= b |
| Arabidopsis thaliana | 3203..6110 | /*tag= c |
| Arabidopsis thaliana | 3203..6110 | /*tag= d |
| Arabidopsis thaliana | 3203..6110 | /*tag= e |
| Arabidopsis thaliana | 3203..6110 | /*tag= f |
| Arabidopsis thaliana | 3203..6110 | /*tag= g |
| Arabidopsis thaliana | 3203..6110 | /*tag= h |
| Arabidopsis thaliana | 3203..6110 | /*tag= i |
| Arabidopsis thaliana | 3203..6110 | /*tag= j |
| Arabidopsis thaliana | 3203..6110 | /*tag= k |
| Arabidopsis thaliana | 3203..6110 | /*tag= l |
| Arabidopsis thaliana | 3203..6110 | /*tag= m |
| Arabidopsis thaliana | 3203..6110 | /*tag= n |
| Arabidopsis thaliana | 3203..6110 | /*tag= o |
| Arabidopsis thaliana | 3203..6110 | /*tag= p |
| Arabidopsis thaliana | 3203..6110 | /*tag= q |
| Arabidopsis thaliana | 3203..6110 | /*tag= r |
| Arabidopsis thaliana | 3203..6110 | /*tag= s |
| Arabidopsis thaliana | 3203..6110 | /*tag= t |
| Arabidopsis thaliana | 3203..6110 | /*tag= u |
| Arabidopsis thaliana | 3203..6110 | /*tag= v |
| Arabidopsis thaliana | 3203..6110 | /*tag= w |
| Arabidopsis thaliana | 3203..6110 | /*tag= x |
| Arabidopsis thaliana | 3203..6110 | /*tag= y |
| Arabidopsis thaliana | 3203..6110 | /*tag= z |

| FT | Intron | 4067..4164 |
|----------------------------|--|------------|
| FT | exon | /*tag- i |
| FT | exon | 4165..4479 |
| FT | Intron | /*tag- j |
| FT | exon | 4480..4631 |
| FT | exon | /*tag- k |
| FT | Intron | 4632..4724 |
| FT | exon | /*tag- l |
| FT | exon | 4725..4815 |
| FT | exon | /*tag- m |
| FT | Intron | 4816..4894 |
| FT | exon | /*tag- n |
| FT | exon | 4895..5000 |
| FT | exon | /*tag- o |
| FT | Intron | 5001..5110 |
| FT | exon | /*tag- p |
| FT | Intron | 5111..5864 |
| FT | exon | /*tag- q |
| FT | exon | 5865..6110 |
| FT | 3'UTR | /*tag- r |
| FT | | 6011..6468 |
| FT | | /*tag- s |
| XX | MO200047715-A2. | |
| PN | 17-AUG-2000. | |
| XX | 11-FEB-2000; 2000MO-US03820. | |
| XX | 11-FEB-1999; 9905-01196557. | |
| PR | 11-FEB-1999; 9905-0119658. | |
| PA | (ARIZ-) ARIZONA BOARD OF REGENTS. | |
| XX | Aspiroz R, Choe S, Feldmann KA; | |
| PI | WPI; 2000-549142/50. | |
| DR | P-PSDB: AAB07921. | |
| XX | New isolated dwf4 polynucleotide useful for altering the phenotype of | |
| PT | plants, for diagnostic assays and in the production of antibodies - | |
| PS | Claim 3; Fig 10A-G; 113pp: English. | |
| XX | The present sequence encodes a DWF4 polypeptide. The polypeptide is a | |
| CC | cyclochrome P450 enzyme that mediates multiple steps in synthesis of | |
| CC | brassinosteroids. Specifically, it mediates multiple | |
| CC | 22alpha-hydroxylation steps in brassinosteroid biosynthesis. The DWF4 | |
| CC | polynucleotide is used for altering the phenotype of a plant. DWF4 | |
| CC | plants display a dramatic reduction in the length of different organs, | |
| CC | and this size reduction is attributable to a defect in cell elongation | |
| CC | The DWF4 polynucleotides and polypeptides can be used in diagnostic | |
| CC | assays and to generate antibodies, which can be used to produce | |
| CC | immunogenic compositions. | |
| XX | Sequence 6888 BP; 2294 A; 1010 C; 1193 G; 2391 T; 0 other; | |
| SO | | |
| Query Match | 100.0%; Score 3889; DB 21; Length 6888; | |
| Best Local Similarity | 100.0%; Pred. No. 0; | |
| Matches 3889; Conservative | 0; Mismatches 0; Indels 0; Gaps | |
| OY | 1 atgaagctcatgtgttaagcttaataagaagaagattatatttaatttaagcat 60 | |
| DB | 3000 atgaagctcatgtgttaagcttaataagaagaattattatatttaattgaagcat 3059 | |
| OY | 61 gatacaattatatttcttactctctttaaaccctcttcaacaagaagctccctt 120 | |
| DB | 3060 gatacaattatatttcttactctctttaaaccctcttcaacaagaagctccctt 3119 | |
| OY | 121 tttagtgaagtcgcgaatcccaattcttaagaacaaagccattagaagaagaagtga 180 | |
| DB | 3120 tttagtgaagtcgcgaatcccaattcttaagaacaaagccattagaagaagaagaagtga 3179 | |

[illegible]

QY 1261 agagatgaacagagatgaatgaagtgatcatgtgttagaacaacaagacagacatga 1320
|||||
Db 4260 agagatgaacagagatgaatgaagtgatcatgtgttagaacaacaagacagacatga 4319
QY 1321 tctttggagatgggttttgaacaatctgaaattatcgacggagacaattctcgatctcat 1380
|||||
Db 4320 tctttggagatgggttttgaacaatctgaaattatcgacggagacaattctcgatctcat 4379
QY 1381 tcttagttgttatgttcgcgagacatgaaacttctctgttagccattgtctcgatctctt 1440
|||||
Db 4380 tcttagttgttatgttcgcgagacatgaaacttctctgttagccattgtctcgatctctt 4439
QY 1441 ctctctgcaagcttgccttaaaagccgttgaagagcttgaaggatataatgaacagca 1500
|||||
Db 4440 ctctctgcaagcttgccttaaaagccgttgaagagcttgaaggatataatgaacagca 4499
QY 1501 caagttaatctctaccaaattgttacgtattatataaagttattatagaattatctctat 1560
|||||
Db 4500 caagttaatctctaccaaattgttacgtattatataaagttattatagaattatctctat 4559
QY 1561 gaataacgatagaanaaagtgtgatatttaattatgttcgaactaatttatgttatgtatt 1620
|||||
Db 4560 gaataacgatagaanaaagtgtgatatttaattatgttcgaactaatttatgttatgtatt 4619
QY 1621 atactttgaaagaaagacatcttgagatcgcgagggccaagaaagaaactaagagatga 1680
|||||
Db 4620 atactttgaaagaaagacatcttgagatcgcgagggccaagaaagaaactaagagatga 4679
QY 1681 gaattaaattggagatgatacaagaanaattggaacttactcaatgtgtatgttactatcat 1740
|||||
Db 4680 gaattaaattggagatgatacaagaanaattggaacttactcaatgtgtatgttactatcat 4739
QY 1741 tctcaatttatatttatgttcatatgattatgattatgataagaacaaatttatgtattttt 1800
|||||
Db 4740 tctcaatttatatttatgttcatatgattatgattatgataagaacaaatttatgtattttt 4799
QY 1801 ttgtgtgtgttgaaggtatataaattgaactcttcgatctggaaatgtatgtatgtttt 1860
|||||
Db 4800 ttgtgtgtgttgaaggtatataaattgaactcttcgatctggaaatgtatgtatgtttt 4859
QY 1861 gcatcgcaaaagcactcaaaagtgttcggtacaaaggttaaaaactttagctacaaaatttt 1920
|||||
Db 4860 gcatcgcaaaagcactcaaaagtgttcggtacaaaggttaaaaactttagctacaaaatttt 4919
QY 1921 aaataatgaataatcggaatatgtgaattcttatgttgatgaataaataatgaataatttaca 1980
|||||
Db 4920 aaataatgaataatcggaatatgtgaattcttatgttgatgaataaataatgaataatttaca 4979
QY 1981 ttcttcaatgttgaanaaagagatacgatacctcctagcggtggtgaaggtgttaacgggtga 2040
|||||
Db 4980 ttcttcaatgttgaanaaagagatacgatacctcctagcggtggtggaaggtgttaacgggtga 5039
QY 2041 tctcaagccgtatacatcttgataattctcgtttatgaccaacctaactctctttaaacttctga 2100
|||||
Db 5040 tctcaagccgtatacatcttgataattctcgtttatgaccaacctaactctctttaaacttctga 5099
QY 2101 gatggcaacaggtataataaagttctctcgtttaactatcgaaaattgaagtgatattt 2160
|||||
Db 5100 gatggcaacaggtataataaagttctctcgtttaactatcgaaaattgaagtgatattt 5159
QY 2161 ttctacatctgacatgataagatacgtctcctagtgatattaccatactatagatactatc 2220
|||||
Db 5160 ttctacatctgacatgataagatacgtctcctagtgatattaccatactatagatactatc 5219
QY 2221 gagaactatcaatctgcaaaaacttttatatatactatcttcaagttaagatcttaac 2280
|||||
Db 5220 gagaactatcaatctgcaaaaacttttatatatactatcttcaagttaagatcttaac 5279
QY 2281 acgtcatatgacatctgacatgataagacataataataaataagagagagaaga 2340
|||||
Db 5280 acgtcatatgacatctgacatgataagacataataataaataagagagagaaga 5339
QY 2341 cgtgtgtgtgaagtgatcgtgaagaacaattatagtagagtggtatgtctttaaagac 2400

|||||
Db 5340 cgtgtgtgtgaagtgatcgtgaagaacaattatagtagagtggtatgtctttaaagac 5399
QY 2401 gtagagctgacctaaatatcttataatcgttgaacggttgatttatattatagtcacgctt 2460
|||||
Db 5400 gtagagctgacctaaatatcttataatcgttgaacggttgatttatattatagtcacgctt 5459
QY 2461 tgatacaattaaagtttgaagagatgatacagcgctttagaaggtatctcaacgctcc 2520
|||||
Db 5460 tgatacaattaaagtttgaagagatgatacagcgctttagaaggtatctcaacgctcc 5519
QY 2521 gatgactgtacgttgcgttagatcttcgcaacggttgaacggttttlaactatgaattata 2580
|||||
Db 5520 gatgactgtacgttgcgttagatcttcgcaacggttgaacggttttlaactatgaattata 5579
QY 2581 actctataattataacaactatcaaatcgcgaatttagagaataatactatagatatt 2640
|||||
Db 5580 actctataattataacaactatcaaatcgcgaatttagagaataatactatagatatt 5639
QY 2641 attagattattatagatataacttattgaaatgaatgaatgaatgaatgaatgaatgaat 2700
|||||
Db 5640 attagattattatagatataacttattgaaatgaatgaatgaatgaatgaatgaatgaat 5699
QY 2701 atagtagtgagggaaggtgaagagtggtgagagatgattatgaccccaagctgtgtgt 2760
|||||
Db 5700 atagtagtgagggaaggtgaagagtggtgagagatgattatgaccccaagctgtgtgt 5759
QY 2761 gccacaagcagctgttct 2820
|||||
Db 5760 gccacaagcagctgttct 5819
QY 2821 ttgtattataaactcgttttgcgttttttttttttttttttttttttttttttttttttt 2880
|||||
Db 5820 ttgtattataaactcgttttgcgttttttttttttttttttttttttttttttttttttt 5879
QY 2881 tcatcgtcaggaagtgtagtttttcgacgttggggaacaaactacatgacgcttttgagga 2940
|||||
Db 5880 tcatcgtcaggaagtgtagtttttcgacgttggggaacaaactacatgacgcttttgagga 5939
QY 2941 gggccaagcctatgtgtcgttcagagctgacgaagtgaagaatggcagtgattatcat 3000
|||||
Db 5940 gggccaagcctatgtgtcgttcagagctgacgaagtgaagaatggcagtgattatcat 5999
QY 3001 catctagttcttaaatcttaattggaatgagagaagatgataccaactttgtcttct 3060
|||||
Db 6000 catctagttcttaaatcttaattggaatgagagaagatgataccaactttgtcttct 6059
QY 3061 ttgtttgatttccctaaacggtttgcctataggttttcgattctcgttaaaaaaaa 3120
|||||
Db 6060 ttgtttgatttccctaaacggtttgcctataggttttcgattctcgttaaaaaaaa 6119
QY 3121 aagatgaagaatttttattctctctctctctctctctctctctctctctctctctct 3180
|||||
Db 6120 aagatgaagaatttttattctctctctctctctctctctctctctctctctctctct 6179
QY 3181 cccaatgataataaanaatttggataaataatatttggataatcgttttttaatttgg 3240
|||||
Db 6180 cccaatgataataaanaatttggataaataatatttggataatcgttttttaatttgg 6239
QY 3241 gtttgaanaaaggtttcgaacttcgaaagtgagagatgataatgattgggaagcttaggt 3300
|||||
Db 6240 gtttgaanaaaggtttcgaacttcgaaagtgagagatgataatgattgggaagcttaggt 6299
QY 3301 tgaactttggaacattgtatgagatgtgtgtgattatagtgatgacacataaaact 3360
|||||
Db 6300 tgaactttggaacattgtatgagatgtgtgtgattatagtgatgacacataaaact 6359
QY 3361 taaatgggcttctataagggccaattatataagataataaagaatgagacactttac 3420
|||||
Db 6360 taaatgggcttctataagggccaattatataagataataaagaatgagacactttac 6419
QY 3421 ttgtttttgacgcgaacaaataaattgtcaattcccaacaaagaattatgataa 3480
|||||

Db 6420 ttcglttttgatccgaagcaataacaaatgtcgaatacccaacaagaatattgtaaa 6479
Oy 3481 cactcgtgtgtctcctagtggaatcattggctgagagactgaacatcgacaacagaa 3540
Db 6480 cactcgtgtgtctcctagtggaatcattggctgagagactgaacatcgacaacagaa 6539
Oy 3541 acccttcaattatgatacaacctcctatgacggttccaaacttattcttctatc 3600
Db 6540 acccttcaattatgatacaacctcctatgacggttccaaacttattcttctatc 6599
Oy 3601 ggtgtaccttgacacaaagagttaagtgcaaaagactaaatgatacaatactctc 3660
Db 6600 ggtgtaccttgacacaaagagttaagtgcaaaagactaaatgatacaatactctc 6659
Oy 3661 aggataagaaggttcataaaatggttctcttatttgagaagaagaagagagactctt 3720
Db 6660 aggataagaaggttcataaaatggttctcttatttgagaagaagaagagagactctt 6719
Oy 3721 actgttctctggtccctatctcctttaaataagagaggttcgttttacttctatctc 3780
Db 6720 actgttctctggtccctatctcctttaaataagagaggttcgttttacttctatctc 6779
Oy 3781 atcacctttagagctcctctctctagacgagtaagaatcctcgttaccaagaatggtct 3840
Db 6780 atcacctttagagctcctctctctagacgagtaagaatcctcgttaccaagaatggtct 6839
Oy 3841 catcttttgaagacaggtcttctccaagctcagtcagtcaggccaagctt 3889
Db 6840 catcttttgaagacaggtcttctccaagctcagtcagtcaggccaagctt 6888

RESULT 2

AAFS8252/c

ID AAF58252 standard; DNA; 936 BP.

AAFS8252;

24-APR-2001 (first entry)

Oligonucleotide D1835.

Electron-transfer group: ETM; mismatch; genotyping;

gene expression; ss.

Synthetic.

WO200107665-A2.

01-FEB-2001.

26-JUL-2000; 2000WO-US20476.

26-JUL-1999; 99US-0145695.

17-MAR-2000; 2000US-0190259.

(CLIN-) CLINICAL MICRO SENSORS INC.

Umek RM;

WPI: 2001-159728/16.

Nucleic acids containing electron-transfer group, useful as labels in hybridization assays, e.g. for genotyping, allowing repeat analyses on a single surface

Example 6; Page 127; 159pp; English.

The present invention relates to a composition comprising two nucleic acids each containing an electron-transfer group (ETM) having different redox potentials. The invention is used for electronic detection of nucleic acids, especially of substitutions (mismatches) and single-nucleotide polymorphisms, e.g. for genotyping, monitoring gene expression.

XX SQ Sequence 936 BP; 4 A; 139 C; 10 G; 7 T; 776 other;

Query Match 5.1%; Score 200; DB 22; Length 936;
Best Local Similarity 1.0%; Pred. No. 5,4e-28;
Matches 8; Conservative 552; Mismatches 232; Indels 0; Gaps 0;

Oy 1479 gggtaagaataataaacaagcacaagtaataatcaccacaatgttgaatataataag 1538
Db 732 GGGCGCMNN 733
Oy 1539 ttattatgaattatcattatgaataatacagatgaataaaaglatatataatgtca 1598
Db 732 WWWNN 673
Oy 1599 ctaattatgttttctatattatcttgaagaaagacatcttgatcgagagc 1658
Db 672 WWWNN 613
Oy 1659 caagaaggaactagagagtcagaataatggatgatatcaagaataatgagacttac 1718
Db 612 WWWNN 553
Oy 1719 tcaatgtatgttactatcatctcattatattctatgtlcatatgatattatgata 1778
Db 552 WWWNN 493
Oy 1779 aaccaaaattatgtattttttttgtgtgtgtgtgaaggtataataatgaactctgat 1838
Db 492 WWWNN 433
Oy 1839 tgggaatgtatgttttctgacacgaagcactcaaatgttgcgtacaaagtta 1898
Db 432 WWWNN 373
Oy 1899 aaactttagtacaataattttaataatgaataatcggaataatgaaacttatgtatg 1958
Db 372 WWWNN 313
Oy 1959 aaaaatattaaataatattacattctcctaagtgtgaaaaaagatacatatccctagt 2018
Db 312 WWWNN 253
Oy 2019 gggtagaaggtgtacccgtgatactcagcgtlacatttgataatctcgttatgacca 2078
Db 252 WWWNN 193
Oy 2079 cctaactccttaactccttgagatgacacaggtataataaagttctcgttaact 2138
Db 192 WWWNN 133
Oy 2139 atcgaaaatagtgatagtttttcatcatttgatgaataatgaatgaactcctagtgat 2198
Db 132 WWWNN 73
Oy 2199 tactatcatagatacactacagaaactataatctgcaaaaactttattatatt 2258
Db 72 WWWNN 13
Oy 2259 atcttcaagtt 2270
Db 12 WWWNNNNNNNNNN 1

RESULT 3

AAFS8254/c

ID AAF58254 standard; DNA; 936 BP.

AAFS8254;

24-APR-2001 (first entry)

[illegible]

```

QY 1959 aaaaatttaaataattacatttcttaattgtaaagaaaagagtagatccctagc 2018
Db 312 wwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwww 253
QY 2019 gggctgaagatgttacgcggtgactcagccgtacatttgataattcgtatgacca 2078
Db 252 wwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwww 193
QY 2079 cctatctcttaactcctctggagatggcaacaggaataaagaattctctgtaact 2138
Db 192 wwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwww 133
QY 2139 atcgaaattagtgtagtcttttcaatcatgcatgaaatagatcagctcctagatt 2198
Db 132 wwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwww 73
QY 2199 tacatcatagatactacacagagaacttaatctggcaaaacttltatatt 2258
Db 72 wwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwww 13
QY 2259 atcttcaagtt 2270
Db 12 wwwwwwwwwwwwwww 1

RESULT 4
AAF58257/c
ID AAF58257 standard; DNA; 936 BP.
XX AC
XX AAF58257;
XX DT
XX 24-APR-2001 (first entry)
XX DN
XX Oligonucleotide D1954.
XX KW
XX Electron-transfer group; ETM; mismatch; genotyping;
XX gene expression; ss.
XX OS
XX Synthetic.
XX PN
XX WO200107665-A2.
XX PD
XX 01-FEB-2001.
XX PE
XX 26-JUL-2000; 2000MO-US20476.
XX PR
XX 26-JUL-1999; 9905-0145695.
XX PR
XX 17-MAR-2000; 2000US-0190259.
XX PA
XX (CLIN-) CLINICAL MICRO SENSORS INC.
XX PI
XX Umek RM;
XX WP1; 2001-159728/16.
XX DR
XX Nucleic acids containing electron-transfer group, useful as labels in
XX hybridization assays, e.g. for genotyping, allowing repeat analyses on
XX a single surface.
XX PT
XX Example 6; Page 127; 159pp; English.
XX PS
XX The present invention relates to a composition comprising two nucleic
XX acids each containing an electron-transfer group (ETM) having
XX different redox potentials. The invention is used for electronic
XX detection of nucleic acids, especially of substitutions (mismatches)
XX and single-nucleotide polymorphisms, e.g. for genotyping,
XX monitoring gene expression.
XX SQ
XX Sequence 936 BP; 5 A; 142 C; 7 G; 6 T; 776 other;

Query Match 5.18; Score 200; DB 22; length 936;

```

[illegible][illegible]

OY 1479 gggttaagataatataacagcacaaagttaattactaccaaatgyltacyltatataga 1538
||| :::::::::::::: : : : :::::::::: : : :::: : : :::::::::::

| | |
|----|-------------|
| XX | 01-FEB-2001 |
| PD | |


```
OY 1602 atttatgtttatgtattatattcttgaagaagacatcttgagatcgagggccaa 1661
DB 121 ..... 180
OY 1662 gaaggaactgaagagatcgaaatgaatggaatgattacaagaatgacttactca 1721
DB 181 ..... 240
OY 1722 atgtgtatgtactatcatctcattattatctatgttctatagattatgagaaac 1781
DB 241 ..... 300
OY 1782 caaattatgtatttttttgggtgtgtgaagtttaaatgaaactctgattgg 1841
DB 301 ..... 360
OY 1842 gaaatgtatgtatgttttgcacgcaagacatcaagaatgttcgtacaagaagtaaa 1901
DB 361 ..... 420
OY 1902 cttacgtacaaaatttttaataatgaatccggaatttgaatcttattgattgaa 1961
DB 421 ..... 480
OY 1962 aatatataaataattacattcttcaatgttgaaaaagatacagataccctagtgg 2021
DB 481 ..... 540
OY 2022 tggaaagtgtacccgtgatactcagccgtacatttggataatctcgtatgacaaact 2081
DB 541 ..... 600
OY 2082 aatccttcaatcccttgagatgacacaggttaataaatttctcctgaatactac 2141
DB 601 ..... 660
OY 2142 gaaaattagttatagtttttcatcatctatgcatgaatagatcactcctgatttac 2201
DB 661 ..... 720
OY 2202 ctactatagatactatacagaaactaatatctcgcacaaactttattattatc 2261
DB 721 ..... 780
OY 2262 ttccaa 2267
DB 781 ..... 786

RESULT 9
AAF58254
ID AAF58254 standard; DNA; 936 BP.
XX
AC AAF58254;
XX
DE 24-APR-2001 (first entry)
XX
XX Oligonucleotide D1875.
XX
KN Electron-transfer group; ETM; mismatch; genotyping;
XX
OS gene expression; ss.
XX
OS Synthetic.
XX
PN WO200107665-A2.
XX
PD 01-FEB-2001.
XX
PF 26-JUL-2000; 2000MO-US20476.
XX
PR 26-JUL-1999; 99US-0145695.
PR 17-MAR-2000; 2000US-0190259.
```

```
XX
PA (CLIN-) CLINICAL MICRO SENSORS INC.
PI Umek RM;
XX WPI; 2001-159728/16.
DR
PT Nucleic acids containing electron-transfer group, useful as labels in
PT hybridization assays, e.g. for genotyping, allowing repeat analyses on
XX a single surface.
XX
PS Example 6; Page 127; 159pp; English.
XX
CC The present invention relates to a composition comprising two nucleic
CC acids each containing an electron-transfer group (ETM) having
CC different redox potentials. The invention is used for electronic
CC detection of nucleic acids, especially of substitutions (mismatches)
CC and single-nucleotide polymorphisms, e.g. for genotyping,
XX monitoring gene expression.
SQ Sequence 936 BP; 4 A; 144 C; 7 G; 5 T; 776 other;
```

Query Match 5.1%; Score 198.8; DB 22; Length 936;
Best Local Similarity 0.6%; Pred. No. 8.9e-28;
Matches 5; Conservative 552; Mismatches 229; Indels 0; Gaps 0;

```
OY 1482 taagataattacaagcaagtaattactaccaaattgttcgtattatagaata 1541
DB 1 ..... 60
OY 1542 ttatagaattattctatagaatacagatgaaaaaagatlatattatgttcaacta 1601
DB 61 ..... 120
OY 1602 atttatgtttatgtattatattcttgaagaagacatcttgagatcgagggccaa 1661
DB 121 ..... 180
OY 1662 gaaggaactgaagagatcgaaatgaatggaatgattacaagaatgacttactca 1721
DB 181 ..... 240
OY 1722 atgtgtatgtactatcatctcattattatctatgttctatagattatgagaaac 1781
DB 241 ..... 300
OY 1782 caaattatgtatttttttgggtgtgtgaagtttaaatgaaactctgattgg 1841
DB 301 ..... 360
OY 1842 gaaatgtatgtatgttttgcacgcaagacatcaagaatgttcgtacaagaagtaaa 1901
DB 361 ..... 420
OY 1902 cttacgtacaaaatttttaataatgaatccggaatttgaatcttattgattgaa 1961
DB 421 ..... 480
OY 1962 aatatataaataattacattcttcaatgttgaaaaagatacagataccctagtgg 2021
DB 481 ..... 540
OY 2022 tggaaagtgtacccgtgatactcagccgtacatttggataatctcgtatgacaaact 2081
DB 541 ..... 600
OY 2082 aatccttcaatcccttgagatgacacaggttaataaatttctcctgaatactac 2141
DB 601 ..... 660
OY 2142 gaaaattagttatagtttttcatcatctatgcatgaatagatcactcctgatttac 2201
DB 661 ..... 720
```

Db 661 www. 720
Qy 2202 ctatctatagatacactagaagaactataatctggaactttttatttattatc 2261
Db 721 www. 780
Qy 2262 ttccaa 2267
Db 781 www. 786

RESULT 10

AAFS8257
ID AAF58257 standard; DNA; 936 BP.

AC AAF58257;

DT 24-APR-2001 (first entry)

DE Oligonucleotide D1954.

KM Electron-transfer group; ETM; mismatch; genotyping;
KW gene expression; ss.

OS Synthetic.

PN WO200107665-A2.

PD 01-FEB-2001.

PF 26-JUL-2000; 2000WO-US20476.

PR 26-JUL-1999; 99US-0145695.

PR 17-MAR-2000; 2000US-0190259.

XX (CLIN-) CLINICAL MICRO SENSORS INC.

PI Umek RM;

DR WPI: 2001-159728/16.

PT Nucleic acids containing electron-transfer group, useful as labels in
PT hybridization assays, e.g. for genotyping, allowing repeat analyses on
PT a single surface

PS Example 6; Page 127; 159pp; English.

CC The present invention relates to a composition comprising two nucleic
CC acids each containing an electron-transfer group (ETM) having
CC different redox potentials. The invention is used for electronic
CC detection of nucleic acids, especially of substitutions (mismatches)
CC and single-nucleotide polymorphisms, e.g. for genotyping,
CC monitoring gene expression.

XX Sequence 936 BP; 5 A; 142 C; 7 G; 6 T; 776 other;

Query Match 5.1%; Score 198.8; DB 22; Length 936;

Best Local Similarity 0.6%; Pred. No. 8.9e-28;

Matches 5; Conservative 552; Mismatches 229; Indels 0; Gaps 0;

Qy 1482 taagataataaacaagaagtaataactacccaattgtacgtattataagta 1541

Db 1 www. 60

Qy 1542 ttatagaattatctatagaataacgatacgaataaagatgataattatgta 1601

Db 61 www. 120

Qy 1602 atttatgttatgtattatactttgaaggaagacatcttgacgcgagggcca 1661

Db 121 www. 180

Qy 1662 gaaggactagagagtcgaactaaatggatgattacaaagaatggacttacct 1721

Db 181 www. 240

Qy 1722 atgtaagttaactacatctcatattatctcatgttcatagattatgataaac 1781

Db 241 www. 300

Qy 1782 caaatattgattttttttgtgtgtgtgtaaggtataaataaaccttgatgg 1841

Db 301 www. 360

Qy 1842 gaaatgtagtatttgcacgcgaagcactcaagaatgttcgtaacaaggtaaa 1901

Db 361 www. 420

Qy 1902 cttaacgtaacaattttaataaagaalccggaataatgaaatcttatgtgaa 1961

Db 421 www. 480

Qy 1962 aatatataaataattacattcttaagtgtgaaataaagatacagataccctagtg 2021

Db 481 www. 540

Qy 2022 tggaaagtgtaacggtacacgcagcgtacatttgataattcgtatgaccaact 2081

Db 541 www. 600

Qy 2082 aatctcttaaccttgagatgcaacaggtaaataaagttctctcgtaactatc 2141

Db 601 www. 660

Qy 2142 gaaatgtagtatttctcatctatgtcatgaatacagatacagctcctagtgat 2201

Db 661 www. 720

Qy 2202 ctatctatagatacactagaagaactataatctggaacttttattatc 2261

Db 721 www. 780

Qy 2262 ttccaa 2267

Db 781 www. 786

RESULT 11

AAFS8259
ID AAF58259 standard; DNA; 936 BP.

AC AAF58259;

DT 24-APR-2001 (first entry)

DE Oligonucleotide D2004.

KM Electron-transfer group; ETM; mismatch; genotyping;
KW gene expression; ss.

OS Synthetic.

PN WO200107665-A2.

PD 01-FEB-2001.

PF 26-JUL-2000; 2000WO-US20476.

PR 26-JUL-1999; 99US-0145695.

PR 17-MAR-2000; 2000US-0190259.

XX (CLIN-) CLINICAL MICRO SENSORS INC.

PI Umek RM;

QY 1722 atgtgtatgttactatcatctcattatttctaatgttcataatgtatttatgtatgaac 1781

```
Db 241 ..... 300
Oy 1782 caaatatgtatgtttttttgtgtgtgtaaggttaataaactcttgatgg 1841
Db 301 ..... 360
Oy 1842 gaaatgtatgtttttgtcgaagcactcaagatgttcgtcaaggttaaa 1901
Db 361 ..... 420
Oy 1902 cttacgtacaaaattttaataatgaatccggaataatgaatctatgtgaa 1961
Db 421 ..... 480
Oy 1962 aatataaataattacattcttaagtgtgaaaaaaggaatccctgtggg 2021
Db 481 ..... 540
Oy 2022 tggaaagtgtacggtgacgcagcgtacatttgataatctcgtatgacaact 2081
Db 541 ..... 600
Oy 2082 aatccttaatccttgagatgcaacaggtaaataaagttctcgttaactc 2141
Db 601 ..... 660
Oy 2142 gaaatgtatgtatgtttttcatcattgtcgaatagatgaactcgaagttaac 2201
Db 661 ..... 720
Oy 2202 ctatctatagatactacagagaactatcatctgcaaaaactttatattatc 2261
Db 721 ..... 780
Oy 2262 ttcaaa 2267
Db 781 ..... 786

RESULT 13
AAFS8255
ID AAF58255 standard; DNA; 938 BP.
XX
AC AAF58255;
XX
DT 24-APR-2001 (first entry)
XX
DE Oligonucleotide D1876.
XX
KM Electron-transfer group; ETM; mismatch; genotyping;
XX
KW gene expression; ss.
XX
OS Synthetic.
XX
PN WO200107665-A2.
XX
PD 01-FEB-2001.
XX
PE 26-JUL-2000; 2000WO-US20476.
XX
PR 26-JUL-1999; 99US-0145695.
XX
PR 17-MAR-2000; 2000US-0190259.
XX
PA (CLIN-) CLINICAL MICRO SENSORS INC.
XX
PI Umek RM;
XX
DR WPI; 2001-159728/16.
XX
PT Nucleic acids containing electron-transfer group, useful as labels in
PT hybridization assays, e.g. for genotyping, allowing repeat analyses on
PT a single surface
```

```
XX
PS Example 6; Page 127; 159pp; English.
XX
CC The present invention relates to a composition comprising two nucleic
CC acids each containing an electron-transfer group (ETM) having
CC different redox potentials. The invention is used for electronic
CC detection of nucleic acids, especially of substitutions (mismatches)
CC and single-nucleotide polymorphisms, e.g. for genotyping,
CC monitoring gene expression.
XX
SQ Sequence 938 BP; 4 A; 144 C; 9 G; 5 T; 776 other;

Query Match 5.1%; Score 198.8; DB 22; Length 938;
Best Local Similarity 0.6%; Pred. No. 8.9e-28;
Matches 5; Conservative 552; Mismatches 229; Indels 0; Gaps 0;

Oy 1482 taagataattataacagacaagttaattactaccaaattgtacgtatataagta 1541
Db 1 ..... 60
Oy 1542 ttatgaattatctcattagaataacgatgaaaaagtatcatattatgcacta 1601
Db 61 ..... 120
Oy 1602 atttatgttatgtattatatactttgaaggaagacatcttgatcgagggccaa 1661
Db 121 ..... 180
Oy 1662 gaagaaactagagagatcagaataattggatgattcaagaataatgacttactca 1721
Db 181 ..... 240
Oy 1722 atgtatgtactatcattccattattatctcattgttcacatgatattagatgaac 1781
Db 241 ..... 300
Oy 1782 caaatatgtatgtttttttgtgtgtgtaaggttaataaactcttgatgg 1841
Db 301 ..... 360
Oy 1842 gaaatgtatgtttttgtcgaagcactcaagatgttcgtcaaggttaaa 1901
Db 361 ..... 420
Oy 1902 cttacgtacaaaattttaataatgaatccggaataatgaatctatgtgaa 1961
Db 421 ..... 480
Oy 1962 aatataaataattacattcttaagtgtgaaaaaaggaatccctagtggg 2021
Db 481 ..... 540
Oy 2022 tggaaagtgtacggtgacgcagcgtacatttgataatctcgttataaccaact 2081
Db 541 ..... 600
Oy 2082 aatccttaatccttgagatgcaacaggtaaataaagttctcgttataactc 2141
Db 601 ..... 660
Oy 2142 gaaatgtatgtatgtttttcatcattgtcgaatagatgaactcgaagttaac 2201
Db 661 ..... 720
Oy 2202 ctatctatagatactacagagaactatcatctgcaaaaactttatattatc 2261
Db 721 ..... 780
Oy 2262 ttcaaa 2267
Db 781 ..... 786
```


GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 17, 2001, 19:13:08 ; Search time 156 Seconds
(without alignments)
4719.429 Million cell updates/sec

Title: US-09-502-426-1_COPY_3000_6888

Perfect score: 3889

Sequence: 1 atgaagctcattgttagtt.....cctagtcagcgcaagctt 3889

Scoring table: IDENTITY_NUC

Gapop 10.0, Gapext 1.0

Searched: 324599 seqs, 9465562 residues

Total number of hits satisfying chosen parameters: 649198

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents_NA.*

1: /cgn2_6/ptodata/2/ina/5A.COMB.seq:*
2: /cgn2_6/ptodata/2/ina/5B.COMB.seq:*
3: /cgn2_6/ptodata/2/ina/6A.COMB.seq:*
4: /cgn2_6/ptodata/2/ina/6B.COMB.seq:*
5: /cgn2_6/ptodata/2/ina/PCIOS.COMB.seq:*
6: /cgn2_6/ptodata/2/ina/backfillseq1.seq:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description |
|------------|-------|-------------|--------|----|--------------------|
| 1 | 92.8 | 2.4 | 4937 | 2 | US-08-622-166A-3 |
| 2 | 87.6 | 2.3 | 1608 | 2 | US-08-622-166A-1 |
| 3 | 62.6 | 1.6 | 7218 | 1 | US-08-232-463-14 |
| 4 | 54.8 | 1.4 | 837 | 4 | US-08-998-416-288 |
| 5 | 54.2 | 1.4 | 665 | 2 | US-08-883-795A-36 |
| 6 | 54 | 1.4 | 636 | 4 | US-08-998-416-1137 |
| 7 | 51.6 | 1.3 | 1251 | 4 | US-08-972-927-7 |
| 8 | 50.2 | 1.3 | 19124 | 2 | US-08-487-826B-13 |
| 9 | 48.8 | 1.3 | 615 | 2 | US-08-998-416-186 |
| 10 | 47.6 | 1.2 | 665 | 2 | US-08-883-795A-36 |
| 11 | 47 | 1.2 | 1461 | 3 | US-08-722-126A-4 |
| 12 | 47 | 1.2 | 1461 | 5 | PCR-US95-04258-4 |
| 13 | 46.8 | 1.2 | 3401 | 1 | US-08-218-943-4 |
| 14 | 46.4 | 1.2 | 1117 | 4 | US-09-247-373B-33 |
| 15 | 46.2 | 1.2 | 518 | 1 | US-08-485-284A-2 |
| 16 | 45.6 | 1.2 | 5852 | 1 | US-07-867-106-2 |
| 17 | 45.4 | 1.2 | 467 | 2 | US-08-841-349-18 |
| 18 | 45.4 | 1.2 | 6671 | 1 | US-08-280-443-1 |
| 19 | 45.4 | 1.2 | 6671 | 1 | US-08-457-459-1 |
| 20 | 45.4 | 1.2 | 6671 | 1 | US-08-555-678-1 |
| 21 | 45.4 | 1.2 | 1474 | 5 | PCR-US95-02275-1 |
| 22 | 45 | 1.2 | 1641 | 1 | US-08-821-994-64 |
| 23 | 45 | 1.2 | 1641 | 1 | US-08-300-903A-8 |
| 24 | 44.8 | 1.2 | 1066 | 1 | US-08-157-101A-4 |
| 25 | 44.8 | 1.2 | 1134 | 3 | US-09-248-335-29 |
| 26 | 44.8 | 1.2 | 1882 | 4 | US-09-370-253-1 |
| 27 | 44.8 | 1.2 | 2082 | 2 | US-08-785-310A-2 |

| | | | | | | |
|----|------|-----|------|---|-------------------|--------------------|
| 28 | 44.8 | 1.2 | 3701 | 4 | US-08-845-258-10 | Sequence 10, Appl |
| 29 | 44.8 | 1.2 | 3701 | 4 | US-08-990-571-10 | Sequence 10, Appl |
| 30 | 44.6 | 1.1 | 6243 | 2 | US-09-056-075-1 | Sequence 1, Appl |
| 31 | 44.2 | 1.1 | 658 | 4 | US-08-998-416-595 | Sequence 595, App |
| 32 | 44.2 | 1.1 | 1736 | 3 | US-09-182-816-22 | Sequence 22, Appl |
| 33 | 44.2 | 1.1 | 1736 | 3 | US-09-182-816-24 | Sequence 24, Appl |
| 34 | 44.2 | 1.1 | 1736 | 3 | US-09-471-528-22 | Sequence 22, Appl |
| 35 | 44.2 | 1.1 | 1736 | 3 | US-09-471-528-24 | Sequence 24, Appl |
| 36 | 44 | 1.1 | 8220 | 2 | US-08-568-459A-11 | Sequence 11, Appl |
| 37 | 44 | 1.1 | 8220 | 2 | US-08-487-826B-11 | Sequence 11, Appl |
| 38 | 43.8 | 1.1 | 998 | 4 | US-09-122-400B-5 | Sequence 5, Appl |
| 39 | 43.8 | 1.1 | 1850 | 3 | US-08-617-860B-32 | Sequence 32, Appl |
| 40 | 43.8 | 1.1 | 3095 | 6 | 5231168-1 | Patent No. 5231168 |
| 41 | 43.8 | 1.1 | 4098 | 2 | US-08-605-106-4 | Sequence 4, Appl |
| 42 | 43.6 | 1.1 | 1098 | 3 | US-09-248-335-35 | Sequence 35, Appl |
| 43 | 43.6 | 1.1 | 1534 | 1 | US-08-300-903A-6 | Sequence 6, Appl |
| 44 | 43.6 | 1.1 | 2550 | 6 | 5258287-23 | Patent No. 5258287 |
| 45 | 43.6 | 1.1 | 2671 | 6 | 5168051-9 | Patent No. 5168051 |

ALIGNMENTS

RESULT 1
US-08-622-166A-3
; Sequence 3, Application US/08622166A
; Patent No. 5952545
; GENERAL INFORMATION:
; APPLICANT: KONCZ, CSABA
; APPLICANT: MATYUR, JAIDEEP
; APPLICANT: SZEKERES, MIKLOS
; APPLICANT: ALTMANN, THOMAS
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING CYTOCHROME
; TITLE OF INVENTION: P450-TYPE PROTEINS INVOLVED IN THE BRASSINOSTEROID SYNTHESIS
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Birch, Stewart, Kolasch & Birch, LLP
; STREET: P.O. Box 747
; CITY: Falls Church
; STATE: Virginia
; COUNTRY: USA
; ZIP: 22040-0747
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/622,166A
; FILING DATE: 27-MAR-1996
; CLASSIFICATION: 800
; ATTORNEY/AGENT INFORMATION:
; NAME: SVENSSON, LEONARD R.
; REGISTRATION NUMBER: 30,330
; REFERENCE/DOCKET NUMBER: 0147-0153P
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 205-8000
; TELEFAX: (703) 205-8050
; TELEFAX: 248345
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4937 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Arabidopsis thaliana
; STRAIN: cv. Columbia
; IMMEDIATE SOURCE:
; LIBRARY: lambda gt10

Db 75 TTTAACTATTATTCATATTTTAAATAATTAATTTGATATTTAATACTATTTATA 134
OY 1581 atgataattatgtcctcaatttattgattatgatttaactttgaaagaagaca 1640
Db 135 TAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 194
OY 1641 tcttgatcgcgagggcgaagaagaactagagagtcagaatcaattggagatga 1700
Db 195 TATTTAATGATGATCTTTTGTCTAGTCAAAATTTTAAATTTAGTTATTTAAATATA 254
OY 1701 caagaagaatgacttaactcaatgctgctgctacatactcattcattcattc 1760
Db 255 TTAGATATTATTTATTTCTTTTAAATTAATTAATTAATTAATTAATTAATTA 314
OY 1761 tcatgatttattgattgaacccaattatgatttttttttggctgctggaagtt 1820
Db 315 TTTATTAATTTGTTTAAATAATTAATTAATTAATTAATTAATTAATTAATTA 374
OY 1821 taaagaacctctcgattggaatgctgattggttttgcacgcaagcactcaag 1880
Db 375 TATTTAATTAATTTATTTTATTAATTAATTAATTTTAAATTTATGTTGATTTA 434
OY 1881 atgtcggtaacaaggtaaacttaccgtaacaattttaaataatgaactcggaa 1940
Db 435 TTAATTAATCTTTTAAAGATTAATTAATTAATTAATTAATTAATTAATTTCTTATTA 494
OY 1941 ttgaactcttattgctgaaataatltaaataattacatttcttaattgctgga 2000
Db 495 TT-AATTTTAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 553
OY 2001 ggaacgataccctagt 2018
Db 554 ATTAATTAATTAATTAAT 571

RESULT 7
US-08-972-927-7/c
Sequence 7, Application US/08972927
Patent No. 6166290
GENERAL INFORMATION:
APPLICANT: Rea, Philip A
APPLICANT: Lu, Yu-Ping
APPLICANT: Li, Ze-Sheng
TITLE OF INVENTION: GLUTATHIONE-S-CONJUGATE TRANSPORT IN
NUMBER OF SEQUENCES: 20
CORRESPONDENCE ADDRESS:
ADDRESSEE: PANITCH SCHWARZE JACOBS & NADEL, P.C.
STREET: One Commerce Square, 2005 Market Street, 22nd
STREET: Floor
CITY: Philadelphia
STATE: Pennsylvania
COUNTRY: US
ZIP: 19103-7086
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/972,927
FILING DATE: 18-NOV-1997
CLASSIFICATION: 800
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 60/031,040
FILING DATE: 18-NOV-1996
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 60/061,328
FILING DATE: 08-OCT-1997
ATTORNEY/AGENT INFORMATION:
NAME: Doyle Leary Ph.D., Kathryn
REGISTRATION NUMBER: 36,317

REFERENCE/DOCKET NUMBER: 9596-1202
TELECOMMUNICATION INFORMATION:
TELEPHONE: 215-965-1284
TELEFAX: 215-967-2991
TELEX: 831-494
INFORMATION FOR SEQ ID NO: 7:
SEQUENCE CHARACTERISTICS:
LENGTH: 1251 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-08-972-927-7

Query Match 1.3%; Score 51.6; DB 4; Length 1251;
Best Local Similarity 45.1%; Pred. No. 0.034;
Matches 192; Conservative 0; Mismatches 234; Indels 0; Gaps 0;

OY 746 aagatgttgagagacatcttctgttctgctgcttctgcaacaactctatttc 805
Db 994 AATGGGTAAATTAATCTGTTCTTTTCTATTTAGGCAACTAATTAATTAAGTTGTTA 935
OY 806 tctgctcaagagagccaaaagtttatttatttatttatttatttatttattt 865
Db 934 TTTCTTACAAATATTTTATTAATTAATTAATTAATTAATTAATTAATTAATTA 875
OY 866 tgttatgaactcttgaagttcttaacttttttttttatttattgaacagttacgtt 925
Db 874 CTTTATTTATTTAATGCTTATTCATTAATTTAGTATTTTGCAAAATTAATTAATA 815
OY 926 tctaattggcgaagcatatgatgatgatgctcctgagagaagaagaacgcaataa 985
Db 814 GGTAAATTAATAATGATATTCATTAAGGAAGAGTGGATTAATTAATTCGTCCAAAGTTAA 755
OY 986 gaaagatagtacttcaatgaagaggttctctgctcctcaaatcaccgaac 1045
Db 754 CAACACTTTTCTTTTGTGAGAAAGATTAATTAATTAATTAATTAATTAATTA 695
OY 1046 tgcctatcaaaagctctcaggtacattatttttttctgctgaagtcacaactct 1105
Db 694 AAGTATTTCTACATTAAGTCAATCTCTTTTGACACTAGTATTTTAAATTTAT 635
OY 1106 catttaagtttatttatttatttatttatttatttatttatttatttatttattt 1165
Db 634 TAACTATCAATGAATAATCATGTTCTTTAAAAAACACATGATTTTATAGATGATCATA 575
OY 1166 tcaaga 1171
Db 574 TCACGA 569

RESULT 8
US-08-487-826B-13/c
Sequence 13, Application US/08487826B
Patent No. 5993827
GENERAL INFORMATION:
APPLICANT: Sim, Kim L.
APPLICANT: Chlitis, Chetan
APPLICANT: Miller, Louis H.
APPLICANT: Peterson, David S.
APPLICANT: Su, Xin-zhuan
TITLE OF INVENTION: BINDING DOMAINS FROM PLASMODIUM VIVAX
NUMBER OF SEQUENCES: 45
CORRESPONDENCE ADDRESS:
ADDRESSEE: Knobb, Martens Olson & Bear
STREET: 620 Newport Center Drive 16th Floor
CITY: Newport Beach
STATE: California
COUNTRY: US
ZIP: 92660

MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: PatentIn Release #1.0, Version #1.30 (EPO)
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: PCT/US95/04258
 FILING DATE: 06-APR-1995
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: IL 109257
 FILING DATE: 08-APR-1994
 ATTORNEY/AGENT INFORMATION:
 NAME: BROWDY, Roger L.
 REGISTRATION NUMBER: 25,618
 REFERENCE/DOCKET NUMBER: PCT-1 PCT
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 202-628-5197
 TELEFAX: 202-737-3528
 TELEX: 248633
 INFORMATION FOR SEQ ID NO: 4:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 1461 base pairs
 TYPE: nucleic acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 MOLECULE TYPE: cDNA
 FEATURE:
 NAME/KEY: CDS
 LOCATION: 54..617
 PCT-US95-04258-4

```

Query Match          1.2%: Score 47; DB 5; Length 1461;
Best Local Similarity 58.0%: Pred. No. 0.32;
Matches 83: Conservative 0; Mismatches 60; Indels 0; Gaps 0;

QY 837 ttttatcttattcttgctaaattttctgttatagaactcttagagttcttaacttt 896
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1460 TTTTATTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT 1401

QY 897 ttttttattgaacagcttcacgctttatctcaagcggaagcctatbaagctatgat 956
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1400 TTTTATTTTACAGAAAGGTAATTTTATTTTAATTTGGACTCTTTTAATTTT7TAGGCAA 1341

QY 957 cctggaagaagaagaacagagca 979
      ||||| ||||| ||||| |||||
Db 1340 ATTAGGATAGATATGATGATGTA 1318

RESULT 13
US-08-218-943-4
; Sequence 4, Application US/08218943
; Patent No. 5614193
; GENERAL INFORMATION:
; APPLICANT: Schmaljohn, Connie S.
; APPLICANT: McClain, David J.
; APPLICANT: Dalrymple, Joel J.
; APPLICANT: Dalrymple, Lonnie
; TITLE OF INVENTION: HANTAVIRUS VACCINE
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 3000 K Street, N.W., Suite 500
; CITY: Washington, D.C.
; COUNTRY: USA
; ZIP: 20007-5109
COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/218,943
; FILING DATE: 28-MAR-1994

```

PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/799,479
FILING DATE: 14-NOV-1991
ATTORNEY/AGENT INFORMATION:
NAME: BENT, Stephen A.
REGISTRATION NUMBER: 29,768
REFERENCE/DOCKET NUMBER: 71007/118/USGO
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202)672-5300
TELEFAX: (202)672-5399
TELEX: 904136
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 3401 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-08-218-943-4

Query Match 1.2%; Score 46.8; DB 1; Length 3401;
Best Local Similarity 50.4%; Pred. No. 0.44;
Matches 114; Conservative 0; Mismatches 112; Indels 0; Gaps 0;

Oy 845 tttttttgctaattttttgtttgtaaacctttagagttcttaactttttttt 904
Db 561 ttttttcaagcgtggagctttgattgacagtttttcattgattgtttttgttc 640
Oy 905 aattgaacagttacgttaacttaataatgacgaacataatgaatgaatcctcgaga 964
Db 641 agttaaggaacttattttaaatttttgacacagtttaagaanaactttgatcaacatg 700
Oy 965 agaaagaaacagagcaatgaagaagatgtaactttcaatgaagaagattgtctgcg 1024
Db 701 caatgatcagaggaatgaagtcgaagatatttattgttattgttggaggaactctgc 760
Oy 1025 tcccttaactaccaggaactgctatcaataaagctcttcagta 1070
Db 761 accattattattgtccacacacttgattgatttcagatccatgacgaaga 806

RESULT 14
US-09-247-373B-33/C
Sequence 33, Application US/09247373B
Patent No. 6168854
GENERAL INFORMATION:
APPLICANT: MCGONIGLE, BRIAN
APPLICANT: O'KEEFE, DANIEL
TITLE OF INVENTION: SOYBEAN GLUTATHIONE-S-TRANSFERASE ENZYMES
FILE REFERENCE: C1-1108-A
CURRENT APPLICATION NUMBER: US/09/247,373B
CURRENT FILING DATE: 1999-02-10
PRIOR APPLICATION NUMBER: 08/924,747
PRIOR FILING DATE: 1997-09-05
NUMBER OF SEQ ID NOS: 56
SOFTWARE: Microsoft Office 97
SEQ ID NO 33
LENGTH: 1117
TYPE: DNA
ORGANISM: SOYBEAN
FEATURE:
NAME/KEY: unsure
LOCATION: (1101)
OTHER INFORMATION: M-A OR C
NAME/KEY: unsure
LOCATION: (1104)
OTHER INFORMATION: M-A OR C
NAME/KEY: unsure
LOCATION: (1116)
OTHER INFORMATION: M-G OR A OR T OR C
US-09-247-373B-33

Query Match 1.2%; Score 46.4; DB 4; Length 1117;
Best Local Similarity 65.4%; Pred. No. 0.39;
Matches 68; Conservative 0; Mismatches 36; Indels 0; Gaps 0;

Oy 831 ttttaatttacttcttaatttgctaattttttgtttacgaatccttaagtttcta 890
Db 1074 ttt 1015
Oy 891 acttttttttttactgaacagttacgtttaacttaatgc 934
Db 1014 tttttttttttttttttttttttaaagaattgaattgattttgtcc 971

RESULT 15
US-08-485-284A-2
Sequence 2, Application US/08485284A
Patent No. 5750372
GENERAL INFORMATION:

APPLICANT: SAKAI, YASUYOSHI
APPLICANT: TANI, YOSHIKI
APPLICANT: SHIBANO, YUJI
APPLICANT: KONDO, HIROTO
APPLICANT: HATANAKA, HARUYO
TITLE OF INVENTION: NOVEL VECTOR HAVING PROMOTER THAT IS
INDUCIBLE BY METHANOL AND/OR GLYCEROL
NUMBER OF SEQUENCES: 18
CORRESPONDENCE ADDRESS:
ADDRESS: CUSHMAN DARBY & CUSHMAN, L.L.P.
STREET: 1100 New York Avenue, N.W.
CITY: Washington
STATE: D. C.

COUNTRY: U.S.A.
ZIP: 20005-3918
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/485,284A
FILING DATE: 07-JUN-1995
CLASSIFICATION: A35
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 43361/1992
FILING DATE: 28-FEB-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/025,416
FILING DATE: 01-MAR-1993
ATTORNEY/AGENT INFORMATION:
NAME: WHITE JR., PAUL E.
REGISTRATION NUMBER: 32,011
REFERENCE/DOCKET NUMBER: 217755/FPS382090S
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 861-3000
TELEFAX: (202) 822-0944
TELEX: 6714627 CUSH

INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 518 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: unknown
MOLECULE TYPE: genomic DNA
US-08-485-284A-2

Query Match 1.2%; Score 46.2; DB 1; Length 518;
Best Local Similarity 54.4%; Pred. No. 0.35;
Matches 93; Conservative 0; Mismatches 78; Indels 0; Gaps 0;

Oy 3048 attgtcttccttctgtgatttcttaacggttgcctataagggttcctgattct 3107
||||| ||||| | ||||| ||| ||||| || | |||||

```

Db      54  ATTTTTCATCAATTTATTTAATTTTCACTTTTAAATCTTGATATGAT 113
Oy      3108 gtaaaaaaaagatgaaagtatttattcctcttcttttttgataatttaa 3167
Db      114  ATGATATGATATGATTTTAGTCTTGCTGTGTTTTTTTTTTTCAACATTTCTT 173
Oy      3168 atcatttcttgcccaatgatalataaaaaatttgataaataatattatt 3218
Db      174  TTAATGACTTTATACCAAAATTTCAAAATTTCCAAAAAACAAT 224

```

Search completed: October 17, 2001, 19:14:07
 Job time: 15722 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 17, 2001, 16:25:00 ; Search time 5433.22 Seconds
(without alignments)
6766.181 Million cell updates/sec

Title: US-09-502-426-1_COPY_3000_6888
Perfect score: 3889
Sequence: 1 atgagctcattggttagtt.....cctagtcagcgccaagctt 3889

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 10228115 segs, 4726426750 residues

Total number of hits satisfying chosen parameters: 20456230

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:*
1: gb_est1:*
2: gb_est2:*
3: gb_est3:*
4: gb_est4:*
5: gb_est5:*
6: gb_est6:*
7: gb_est7:*
8: gb_est8:*
9: gb_est9:*
10: gb_est10:*
11: gb_est11:*
12: gb_est12:*
13: gb_est13:*
14: gb_est14:*
15: gb_est15:*
16: gb_est16:*
17: gb_est17:*
18: gb_est18:*
19: gb_est19:*
20: gb_est20:*
21: gb_est21:*
22: gb_est22:*
23: gb_est23:*
24: gb_est24:*
25: gb_est33:*
26: gb_est34:*
27: gb_est35:*
28: gb_est36:*
29: gb_est37:*
30: gb_est38:*
31: gb_est39:*
32: gb_est40:*
33: em_estba:*
34: em_estfun:*
35: em_esthum1:*
36: em_esthum2:*
37: em_esthum3:*
38: em_esthum4:*
39: em_esthum5:*
40: em_esthum6:*
41: em_esthum7:*
42: em_esthum8:*
43: em_esthum9:*

44: em_esthum10:*
45: em_esthum11:*
46: em_esthum12:*
47: em_esthum13:*
48: em_esthum14:*
49: em_esthum15:*
50: em_esthum16:*
51: em_esthum17:*
52: em_esthum18:*
53: em_esthum19:*
54: em_esthum20:*
55: em_esthum21:*
56: em_esthum22:*
57: em_esthum23:*
58: em_esthum24:*
59: em_esthum25:*
60: em_esthum26:*
61: em_esthum27:*
62: em_esthum28:*
63: em_est11:*
64: em_est12:*
65: em_est13:*
66: em_est14:*
67: em_est15:*
68: em_estom1:*
69: em_estom2:*
70: em_estov1:*
71: em_estov2:*
72: em_estp11:*
73: em_estp12:*
74: em_estp13:*
75: em_estp14:*
76: em_estp15:*
77: em_estp16:*
78: em_estp17:*
79: em_estp18:*
80: em_estp19:*
81: em_estp110:*
82: em_estro1:*
83: em_estro2:*
84: em_estro3:*
85: em_estro4:*
86: em_estro5:*
87: em_estro6:*
88: em_estro7:*
89: em_estro8:*
90: em_estro9:*
91: em_estro10:*
92: em_estro11:*
93: em_estro12:*
94: em_estro13:*
95: em_estro14:*
96: em_estro15:*
97: em_estro16:*
98: em_estro17:*
99: em_estro18:*
100: em_estro19:*
101: em_estro20:*
102: gb_est25:*
103: gb_est26:*
104: gb_est27:*
105: gb_est28:*
106: gb_est29:*
107: gb_est30:*
108: gb_est31:*
109: gb_est32:*
110: gb_est41:*
111: gb_est42:*
112: gb_est43:*
113: gb_est44:*
114: gb_est45:*
115: gb_est46:*
116: gb_est47:*

117: gb_est48:*
118: gb_est49:*
119: gb_est50:*
120: gb_est51:*
121: gb_est52:*
122: gb_est53:*
123: gb_est54:*
124: gb_est55:*
125: gb_est56:*
126: gb_est57:*
127: gb_est58:*
128: gb_est59:*
129: gb_est60:*
130: gb_est61:*
131: gb_est62:*
132: gb_est63:*
133: gb_est64:*
134: gb_est65:*
135: gb_est66:*
136: gb_est67:*
137: gb_est68:*
138: gb_est69:*
139: gb_est70:*
140: gb_est71:*
141: gb_est72:*
142: gb_est73:*
143: gb_est74:*
144: gb_est75:*
145: gb_est76:*
146: gb_est77:*
147: gb_est78:*
148: gb_est79:*
149: gb_est80:*
150: gb_est81:*
151: gb_est82:*
152: gb_est83:*
153: gb_est84:*
154: gb_est85:*
155: gb_est86:*
156: gb_est87:*
157: gb_est88:*
158: gb_est89:*
159: gb_est90:*
160: gb_est91:*
161: gb_est92:*
162: gb_est93:*
163: gb_est94:*
164: gb_est95:*
165: gb_est96:*
166: gb_est97:*
167: gb_est98:*
168: gb_est99:*
169: gb_est100:*
170: gb_est101:*
171: gb_est102:*
172: gb_est103:*
173: gb_est104:*
174: gb_est105:*
175: gb_est106:*
176: gb_est107:*
177: gb_est108:*
178: gb_est109:*
179: gb_est110:*
180: gb_est111:*
181: gb_est112:*
182: gb_est113:*
183: gb_est114:*
184: gb_est115:*
185: gb_est116:*
186: gb_est117:*
187: gb_est118:*
188: gb_est119:*
189: gb_est120:*

190: gb_est121:*
191: gb_est122:*
192: gb_est123:*
193: gb_est124:*
194: gb_est125:*
195: gb_est126:*
196: gb_est127:*
197: gb_est128:*
198: gb_est129:*
199: gb_est130:*
200: gb_est131:*
201: gb_est132:*
202: gb_est133:*
203: gb_est134:*
204: gb_est135:*
205: gb_est136:*
206: gb_est137:*
207: gb_est138:*
208: gb_est139:*
209: gb_est140:*
210: gb_est141:*
211: gb_est142:*
212: gb_est143:*
213: gb_est144:*
214: gb_est145:*
215: gb_est146:*
216: gb_est147:*
217: gb_est148:*
218: gb_est149:*
219: gb_est150:*
220: gb_est151:*
221: gb_est152:*
222: gb_est153:*
223: gb_est154:*
224: gb_est155:*
225: gb_est156:*
226: gb_est157:*
227: gb_est158:*
228: gb_est159:*
229: gb_est160:*
230: gb_est161:*
231: gb_est162:*
232: gb_est163:*
233: gb_est164:*
234: gb_est165:*
235: gb_est166:*
236: gb_est167:*
237: gb_est168:*
238: gb_est169:*
239: gb_est170:*
240: gb_est171:*
241: gb_est172:*
242: gb_est173:*
243: gb_est174:*
244: gb_est175:*
245: gb_est176:*
246: gb_est177:*
247: gb_est178:*
248: gb_est179:*
249: gb_est180:*
250: gb_est181:*
251: gb_est182:*
252: gb_est183:*
253: gb_est184:*
254: gb_est185:*
255: gb_est186:*
256: gb_est187:*
257: gb_est188:*
258: gb_est189:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description |
|------------|-------|-------------|--------|-------|-------------|
| 1 | 541.4 | 13.9 | 543 | 256 | B77695 |
| 2 | 186.8 | 4.8 | 642 | 166 | BE315722 |
| 3 | 178.6 | 4.6 | 667 | 147 | BF324723 |
| 4 | 171.4 | 4.4 | 533 | 143 | BF050501 |
| 5 | 162 | 4.2 | 522 | 152 | BG316131 |
| 6 | 149 | 3.8 | 477 | 143 | BF052296 |
| 7 | 146 | 3.8 | 589 | 232 | AO689106 |
| 8 | 105.4 | 2.7 | 465 | 173 | BO043191 |
| 9 | 105 | 2.7 | 615 | 115 | AW398669 |
| 10 | 103.6 | 2.7 | 191 | 143 | BF052051 |
| 11 | 103.4 | 2.7 | 574 | 118 | AW615836 |
| 12 | 103.4 | 2.7 | 574 | 118 | AW615967 |
| 13 | 100 | 2.6 | 677 | 153 | BG448552 |
| 14 | 97.2 | 2.5 | 524 | 119 | AW690507 |
| 15 | 95.8 | 2.5 | 461 | 120 | AW759817 |
| 16 | 89.2 | 2.3 | 468 | 113 | AW221826 |
| 17 | 87.6 | 2.3 | 583 | 104 | AI993784 |
| 18 | 87.6 | 2.3 | 686 | 162 | BE039087 |
| 19 | 85.2 | 2.2 | 536 | 159 | N65494 |
| 20 | 84.4 | 2.2 | 533 | 167 | BE450142 |
| 21 | 80.4 | 2.1 | 1101 | 219 | CNS00172 |
| 22 | 79.6 | 2.0 | 1101 | 219 | CNS00396 |
| 23 | 75.2 | 1.9 | 1101 | 219 | CNS00591 |
| 24 | 72.2 | 1.9 | 572 | 11 | AA740068 |
| 25 | 72.2 | 1.9 | 1101 | 219 | CNS0100X |
| 26 | 71.8 | 1.8 | 1101 | 219 | CNS008X3 |
| 27 | 71.6 | 1.8 | 342 | 24 | AI725255 |
| 28 | 71.6 | 1.8 | 391 | 31 | AV526355 |
| 29 | 71.2 | 1.8 | 655 | 151 | BF642437 |
| 30 | 70 | 1.8 | 1101 | 219 | CNS00591 |
| 31 | 70 | 1.8 | 610 | 119 | AW648875 |
| 32 | 70 | 1.8 | 1092 | 220 | CNS020K7 |
| 33 | 69.6 | 1.8 | 604 | 11 | AA739879 |
| 34 | 69.4 | 1.8 | 786 | 219 | CNS009A5 |
| 35 | 69.2 | 1.8 | 579 | 167 | BE449595 |
| 36 | 69.2 | 1.8 | 1068 | 219 | CNS00591 |
| 37 | 68.8 | 1.8 | 362 | 31 | AV526026 |
| 38 | 68.8 | 1.8 | 1101 | 219 | CNS00591 |
| 39 | 68 | 1.7 | 928 | 219 | CNS00DKY |
| 40 | 68 | 1.7 | 1101 | 219 | CNS003DQ |
| 41 | 67.8 | 1.7 | 508 | 151 | BF595940 |
| 42 | 67.8 | 1.7 | 524 | 7 | AA404883 |
| 43 | 67.6 | 1.7 | 1101 | 219 | CNS000D1 |
| 44 | 67.6 | 1.7 | 1101 | 219 | CNS00E14 |
| 45 | 67.4 | 1.7 | 1101 | 219 | CNS01450 |

ALIGNMENTS

| RESULT | 1 | 543 bp | DNA | GSS | 16-JAN-1998 |
|------------|---|------------|------|--|-------------|
| LOCUS | B77695 | T28K14TF | RAMU | Arabidopsis thaliana genomic clone T28K14, DNA | |
| DEFINITION | Sequence. | | | | |
| ACCESSION | B77695 | GI:2774334 | | | |
| VERSION | B77695.1 | | | | |
| KEYWORDS | GSS. | | | | |
| SOURCE | thale cress. | | | | |
| ORGANISM | Arabidopsis thaliana | | | | |
| REFERENCE | Eukaryota: Viridiplantae: Streptophyta: Embryophyta: Tracheophyta; | | | | |
| AUTHORS | Rounsley, S.D., Field, C.E., Bass, S., Linher, K., Linher, K., Golden, K., Berry, K., Granger, D., Suh, E., Wible, C., Adams, M.D. and Venter, J.C. | | | | |

TITLE A BAC End Sequence Database for Identifying Minimal Overlaps in Arabidopsis Genomic Sequencing. Update 3
JOURNAL Unpublished (1997)
COMMENT Contact: Steve Rounsley
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: rounsley@tigr.org
Seq primer: M3-21
Class: BAC ends
High quality sequence stop: 543.

FEATURES

source

1.543
/organism="Arabidopsis thaliana"
/strain="Columbia"
/db_xref="taxon:3702"
/clone="T28K14"
/clone.lib="RAMU"
/sex="hermaphrodite"
/note="Vector: BeIOBACII; Site-1: HindIII; Site-2: HindIII
; Produced by Rod Wing"
BASE COUNT 198 a 57 c 95 g 193 t
ORIGIN

Query Match 13.9%; Score 541.4; DB 256; Length 543;
Best Local Similarity 99.8%; Pred. No. 3,6e-92;
Matches 542; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1456 ccctaagccgttgagagcgttgagatgatatataacgacacagttattctac 1515
|||||
DB 1 CCTTAAGCCGTTGAGAGCGTTAGGTAAGTAATTAACACAGTTAATTAAC 60
QY 1516 caaatgttaccgttatataatgattatataacattctatagaataacagtaaa 1575
|||||
DB 61 CAATGTGTAACGTATATATTAATTAATTAATTAATTAATTAATTAATTAAT 120
QY 1576 aaagtatgatatataatgttcacataattatgttattatgattatgaagaa 1635
|||||
DB 121 AAGGTATGATATATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 180
QY 1636 gacacattgagatcgccggggccaaagaagaaactgagagtcagaatattggat 1695
|||||
DB 181 GACCACTTCGATCGCGAGGCGCAAGAGGACCTGAGAGTCAAGATTAATTGGAT 240
QY 1696 gattacaagaanaatgacttactcaatgttgatgttactatcatctctattatc 1755
|||||
DB 241 GATTACAGAGAAATGACCTTACTCAATGTGTTACTATATCTCATTTATTATTC 300
QY 1756 tatgttcataatgattatgataagaaacaaatattgatttttttgggtgtgaa 1815
|||||
DB 301 TATGTTCAATGATGATTAATGATGAACCAAAATTAATGATTTTGGTGATGA 360
QY 1816 ggttataatgaacattcgcgattgggaaatgagttggtttttcattcgcaaacct 1875
|||||
DB 361 GGTATTAATGAACCTTCGATTTGGAAATGATGATTTTTCATTCGCAACACT 420
QY 1876 caaagatgttcgttacaagaagtaaacatttacttacaataattttaaataaaccg 1935
|||||
DB 421 CAAGATGTTTCGTTACAAAGTAAACCTTTACTACAAATTTTAAATTAATTAATTCG 480
QY 1936 gaattatgaatcttattgataaataataataataattacattcttaattgttga 1995
|||||
DB 481 GAATATTGAATCTTATTTGGATGAAAAATTAATTAATTAATTTTCAATTTTAAGTTGA 540
QY 1996 aaa 1998
|||
DB 541 AAA 543

RESULT 2

Db 278 GCAGAGCTCAGACAGGTTCTTCTACAAAGAGGAATGTTGCGATGACATCTCT 337
 Qy 623 aagaatgatgggagatcttgggaatggtgatgctgttctgttgatgacatgcat 682
 Db 338 AGAAGCATGGTGGAATCTAGAAATATGCTTGTCTTAACTAGTGCATGATCAT 397
 Qy 683 agagataatgaagatcttcgcttaactcttcaatgaacgacgcttgaactatcca 742
 Db 398 AGAGACATCGGGTTTATCATCACTTCTTAAAGCCAGCCAGCTCGAAGACACCTTC 457
 Qy 743 cctaaagatgttgaagacatcttgttctgtctgtatcttgcgaacaaactatct 802
 Db 458 TTGAAGAGGTGGAGAGATATCCCTCTTGTGATCTTGAAGCCCAAGATTGCAAT 517
 Qy 803 tctctgtcctcaagagagcgaagcgaagcgt 831
 Db 518 TTCTCAGCCCAAGATGAAGTAAAGATT 546

RESULT 4
 LOCUS BF050501 533 bp mRNA EST 16-OCT-2000
 DEFINITION EST435659 tomato developing/immature green fruit Lycopersicon
 esculentum cDNA clone CLEM18G22 5' sequence, mRNA sequence.
 ACCESSION BF050501
 VERSION BF050501.1 GI:10804397
 KEYWORDS EST.
 SOURCE tomato.
 ORGANISM Lycopersicon esculentum
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 Asteridae; euasterids I; Solanales; Solanaceae; Solanum;
 Lycopersicon.
 REFERENCE 1 (bases 1 to 533)
 AUTHORS Alcalá, J., Vrebalov, J., White, R., van der Hoeven, R.S., Holt, I.E.,
 Liang, F., Hansen, T.S., Craven, M.B., Bowman, C.L., Rønning, C.M.,
 Niemann, W., Fraser, C.M., Martin, G.B., Giovannoni, J.J. and Tanksley,
 S.D.
 TITLE Generation of ESTs from tomato fruit tissue, immature green
 Unpublished (2000)
 JOURNAL Contact: David Frisch
 COMMENT Clemson University Genomics Institute
 Clemson University
 100 Jordan Hall, Clemson, SC 29634, USA
 Tel: 864 656 4366
 Fax: 864 656 4293
 Email: dfrisch@CLEMSON.EDU.
 FEATURES
 source
 1. 533
 /organism="Lycopersicon esculentum"
 /cultivar="T496"
 /db_xref="taxon:4081"
 /clone_lib="CLEM18G22"
 /clone_lib="tomato developing/immature green fruit"
 /tissue_type="fruit"
 /dev_stage="immature green (5-35 days post-anthesis)"
 /lab_host="SOLR"
 /note="Vector: pBluescriptSKmClnadap1; Site_1: EcoRI;
 Site_2: Xho1; Fruit were tagged at 5 dpa (0.5 cm) and
 harvested at 7 day intervals through 35 dpa. Equal masses
 of tissue from each stage were combined (including seeds
 and locules) prior to mRNA isolation."
 BASE COUNT 172 a 75 c 111 g 175 t
 ORIGIN

Query Match 4.4%; Score 171.4; DB 143; Length 533;
 Best Local Similarity 73.1%; Pred. No. 1.9e-22;
 Matches 220; Conservative 0; Mismatches 81; Indels 0; Gaps 0;

Qy 498 ttacagcatatgtaagatatagatgaactgttggagacacacagatcgtatcag 557
 Db 233 TTCTAGATATGGAGAAATTTTCAAGTCAAAATTTGTTGGAGAGCCACAAATAGTTTCAAG 292

Qy 558 ctgatgtgacttaataagatatattacaacgaaggaagcctcttgaatgagt 617
 Db 293 CAGATGACAGGGCTAAACACATATCTCTCAGATGAAGGACATTATTTGATTAAT 352
 Qy 618 atcctagaagatagtggtgacttctgggaatgctgcagatgcttctgttgagaca 677
 Db 353 ATCCAGAGATAGGTGGATGACTTCTGTAAATGCTATGTTAGTTCAAGTTGACAAA 412
 Qy 678 tgcataagatatgagaagatctcgttaactcttcaagtaacgacgcttgaacta 737
 Db 413 TGCATAGACATATGAGAGATGATCTCTGAATTTTGAAGCAATGCTAGGCTAAGGAATC 472
 Qy 738 tctacttaagatgttggagacataacttgttcttcttgatcttctgcaacaaact 797
 Db 473 AACTTTAAGTGAAGTTGAAAGCATATCTTGTCTTGTGCTTGGCTTGGAAACAGATT 532
 Qy 798 c 798
 Db 533 c 533

RESULT 5
 LOCUS BG316131
 DEFINITION BG316131
 ID: Gm-c1032-2967 5' similar to TR:064989 064989 STERIOD
 22-ALPHA-HYDROXYLASE. [1] ;, mRNA sequence.
 ACCESSION BG316131
 VERSION BG316131.1 GI:13125561
 KEYWORDS EST.
 SOURCE soybean.
 ORGANISM Glycine max
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Glycine.
 REFERENCE 1 (bases 1 to 522)
 AUTHORS Shoemaker, R., Kelm, P., Vodkin, L., Erpelting, J., Corryell, V., Khanna,
 A., Bolla, B., Marra, M., Hillier, L., Kucaba, T., Martin, J., Beck, C.,
 Wylie, T., Underwood, K., Steptoe, M., Theising, B., Allen, M., Bowers,
 Y., Person, B., Swaller, T., Gibbons, M., Page, D., Harvey, N., Schurr,
 R., Ritter, E., Kohn, S., Shiu, T., Jackson, Y., Cardenas, M., McCann,
 R., Materslop, R. and Wilson, R.
 TITLE Public Soybean EST Project
 JOURNAL Unpublished (1993)
 COMMENT Contact: Shoemaker R./Public Soybean EST Project
 Public Soybean EST Project
 Washington University School of Medicine
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
 Tel: 314 286 1800
 Fax: 314 286 1810
 Email: est@wustl.edu
 This clone is available through: Genome Systems, Inc. 4633 World
 Parkway Circle St. Louis, Missouri 63134 For further information
 call: (800) 430-0030 or (314) 427-3322 FAX: (888) 919-3324 or (314)
 427-3324 or contact: clones@genomesystems.com or
 info@genomesystems.com web site: www.genomesystems.com
 High quality sequence stop: 388.
 FEATURES
 source
 1. 522
 /organism="Glycine max"
 /db_xref="taxon:3847"
 /clone_lib="GENOME SYSTEMS CLONE ID: Gm-c1032-2967"
 /clone_lib="Gm-c1032"
 /tissue_type="Cotyledons of 8-day-old 'Williams' seedlings"
 /lab_host="DH10B"
 /note="Vector: pBluescript.II XR; Site_1: EcoRI; Site_2:
 XhoI; This cDNA library was constructed from mRNA isolated
 from cotyledons of 8-day-old 'Williams' seedlings which
 were propagated on paper towels with distilled water for
 3 days (etiolated), then greenhouse grown for 5 days in
 potting soil. The cotyledons were flash-frozen in liquid

FEATURES
source
Fax: 864 656 4293
Email: rtwing@clmson.edu
Seq primer: GGAAACAGCTATATCAGCATG
Class: BAC ends
High quality sequence stop: 191
Location/Qualifiers
1..589

/note="Vector: pBI12AC11; Site1: HindIII; Site2: HindIII". Rice is one of two most popular grains in the world. Half of the world population especially those inhabiting highly populated areas of the humid tropics and subtropics, rely on rice as their primary source of carbohydrate. Monocotyledonous rice is a diploid plant (2n=24) with a haploid genome equivalent of 431 Mbp (Armuganathan and Earle, 1991). The relatively small genome of rice, three times larger than that of Arabidopsis, makes it suitable for genomic studies. In order to facilitate positional cloning, physical mapping and genome sequencing of rice, we have constructed a BAC library from *Oryza sativa*, Nipponbare variety. The library contains 36,864 clones with an average insert size of 128.5 kb providing 10.9 haploid genome equivalents. The deep coverage allows the isolation a particular sequence with a probability of 99.9%. Two high density filters, each containing 18,432 clones (doubly spotted), represent the whole library for colony screening."

QY 962 agaagaagaacagagc 978
| | | | | | | | | |
Db 565 GGGAGAGAGACCGGAGC 581

| REFERENCE | AUTHORS | TITLE | JOURNAL | COMMENT |
|--------------------|--|----------------------------|--------------------|---|
| 1 (Pages 1 to 465) | Shoemaker, R., Keim, P., Vodkin, L., Erpelding, J., Corvelli, V., Khamza, A., Bolla, B., Marra, M., Hiller, L., Kuaba, T., Martin, J., Beck, C., Wyle, T., Underwood, K., Steptoe, B., Theising, B., Allen, M., Bowers, Y., Person, B., Swaller, T., Gibbons, M., Page, D., Harvey, N., Schurr, R., Ritter, E., Kohn, S., Shin, T., Jackson, Y., Cardenas, M., McCann, R., Waterson, R. and Wilson, R. | Public Soybean EST Project | Unpublished (1999) | Contact: Shoemaker R/Public Soybean EST Project |

Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@wustl.edu
This clone is available through: Genome Systems, Inc. 4633 World
Parkway Circle St. Louis, Missouri 63134 For further information
call: 1-800-430-0030 or (314) 427-3222 FAX: (888) 919-3324 or (314)
427-3324 OR contact: clones@genomesystems.com OR
InfoGenomesystems.com web site: www.genomesystems.com
High quality sequence stop: 415.

```

FEATURES
SOURCE

Location/Qualifiers
1. 465
/organism="Glycine max"
/db_xref="taxon:3847"
/clone="GENOME SYSTEMS CLONE ID: Gm-cl066-24"
/clone_id="Gm-cl066"
/tissue_type="leaf and shoot tip, salt stressed, 2 week
old seedling"
/lab_host="DH10B"
/note="Vector: pluescript II SK+, Site_1: EcoRI, Site_2:
XhoI; The cDNA library was constructed from mRNA isolated
from unexpanded leaves and the shoot tips of 2 week old
seedling from the cultivar Williams. The 2 week old
seedlings were salt stressed in a solution of 500mM NaCl
for 3 days prior to harvesting. Complementary DNA was
synthesized from mRNA using a primer consisting of a
poly(dT) sequence with a XhoI restriction site. EcoRI
adapters were ligated to the blunt-ended cDNA fragments
followed by XhoI digestion. The cDNA fragments were
directionally cloned into the EcoRI-XhoI restriction site
of the pluescript vector. The ligated cDNA fragments were
transformed into DH10B host cells (GibcoBRL). This library
was constructed in the laboratory of Dr. Randy
Shoemaker."

```

| | | | | |
|---------------------------|--------|------------------|-----------|-------------|
| BASE COUNT | 144 a | 110 c | 87 g | 124 t |
| ORIGIN | | | | |
| Query Match | 2.7% | Score 105.4; | DB 173; | Length 465; |
| Best Local Similarity | 74.3%; | Pred. NO. 5e-10; | | |
| Matches 133; Conservative | 0; | Mismatches 46; | Indels 0; | Gaps 0; |

621 ctgaagatagtcgggattctctgggaaatgctgatctctctt-gttggctgaca-t 678

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 17, 2001, 19:48:49 ; Search time 9769.47 Seconds
(without alignments)
566.812 Million cell updates/sec

Title: US-09-502-426-1_COPY_6111_6468

Perfect score: 358
Sequence: 1 aaaaaaaaaagatgaaagt.....tgccaataccacaacaag 358

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1344157 seqs, 773874588 residues

Total number of hits satisfying chosen parameters: 2688314

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

GenEmbl:*
1: gb_ba1:*
2: gb_ba2:*
3: gb_ba3:*
4: gb_in1:*
5: gb_in2:*
6: gb_in3:*
7: gb_om:*
8: gb_ov:*
9: gb_pat1:*
10: gb_pat2:*
11: gb_ph:*
12: gb_pi1:*
13: gb_pi2:*
14: gb_pi3:*
15: gb_pi4:*
16: em_ba1:*
17: em_ba2:*
18: em_fun:*
19: em_htgo_hum:*
20: em_htgo_inv:*
21: em_htgo_rod:*
22: em_htg_hum1:*
23: em_htg_hum2:*
24: em_htg_hum3:*
25: em_htg_hum4:*
26: em_htg_hum5:*
27: em_htg_hum6:*
28: em_htg_hum7:*
29: em_htg_hum8:*
30: em_htg_inv1:*
31: em_htg_inv2:*
32: em_htg_other:*
33: em_htg_rod:*
34: em_hum1:*
35: em_hum2:*
36: em_hum3:*
37: em_hum4:*
38: em_hum5:*
39: em_hum6:*
40: em_hum7:*
41: em_in:*
42: em_om:*
43: em_or:*

44: em_ov:*
45: em_pat:*
46: em_ph:*
47: em_pi:*
48: em_ro:*
49: em_sts:*
50: em_sy:*
51: em_un:*
52: em_vl:*
53: gb_sts1:*
54: gb_sts2:*
55: gb_sts3:*
56: gb_sy:*
57: gb_un:*
58: gb_vl1:*
59: gb_vl2:*
60: gb_htg1:*
61: gb_htg2:*
62: gb_htg3:*
63: gb_htg4:*
64: gb_htg5:*
65: gb_htg6:*
66: gb_htg7:*
67: gb_htg8:*
68: gb_htg9:*
69: gb_htg10:*
70: gb_htg11:*
71: gb_htg12:*
72: gb_htg13:*
73: gb_htg14:*
74: gb_htg15:*
75: gb_htg16:*
76: gb_htg17:*
77: gb_htg18:*
78: gb_htg19:*
79: gb_htg20:*
80: gb_htg21:*
81: gb_htg22:*
82: gb_htg23:*
83: gb_htg24:*
84: gb_htg25:*
85: gb_pr1:*
86: gb_pr2:*
87: gb_pr3:*
88: gb_pr4:*
89: gb_pr5:*
90: gb_pr6:*
91: gb_pr7:*
92: gb_pr8:*
93: gb_pr9:*
94: gb_ro1:*
95: gb_ro2:*
96: gb_in4:*
97: gb_pr10:*
98: em_ba3:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | % Match | Query Length | DB ID | Description |
|------------|-------|---------|--------------|-------|--------------------|
| 1 | 358 | 100.0 | 4818 | 12 | AF044216 |
| 2 | 356.4 | 99.6 | 84196 | 14 | ATT3A5 |
| 3 | 56.6 | 15.8 | 1147 | 53 | CNS073CX |
| 4 | 55.6 | 15.5 | 169546 | 60 | AC004157 |
| 5 | 55 | 15.4 | 1007 | 53 | CNS06X9S |
| 6 | 54.2 | 15.1 | 204652 | 84 | PFMAL13P6 |
| 7 | 53.6 | 15.0 | 1167 | 53 | CNS07360 |
| 8 | 53.4 | 14.9 | 973 | 53 | CNS071LE |
| | | | | | AL425064 clone BAO |

| gene | TATA_signal | mRNA | CDS | BASE COUNT | ORIGIN |
|--|--|--|-----|------------|--------|
| /map="8 cm from ngac" | /gene="DMF4" | /gene="DMF4" | | 1566 a | |
| 990. .4398 | /note="dwarf4" | /allele="dwf4-1; dwf4-2; dwf4-3; dwf4-4" | | 727 c | |
| /gene="DMF4" | 990. 1055 | /gene="DMF4" | | 888 g | |
| /note="dwarf4" | join(<1133. .1353,1434. .1758,1844. .1996,2095. .2409, | join(1133. .1353,1434. .1758,1844. .1996,2095. .2409, | | 1637 t | |
| /allele="dwf4-1; dwf4-2; dwf4-3; dwf4-4" | 2562. .2654,2746. .2824,2931. .3040,3795. .4398) | 2562. .2654,2746. .2824,2931. .3040,3795. .4040) | | | |
| 990. 1055 | /gene="DMF4" | /gene="DMF4" | | | |
| /note="dwarf4" | /product="steroid 22-alpha-hydroxylase" | /product="steroid 22-alpha-hydroxylase" | | | |
| /allele="dwf4-1; dwf4-2; dwf4-3; dwf4-4" | join(1133. .1353,1434. .1758,1844. .1996,2095. .2409, | join(1133. .1353,1434. .1758,1844. .1996,2095. .2409, | | | |
| 990. 1055 | 2562. .2654,2746. .2824,2931. .3040,3795. .4040) | 2562. .2654,2746. .2824,2931. .3040,3795. .4040) | | | |
| /gene="DMF4" | /function="brassinosteroid biosynthetic enzyme" | /function="brassinosteroid biosynthetic enzyme" | | | |
| /note="dwarf4" | /note="member of the cytochrome P450 superfamily; CYP90B1" | /note="member of the cytochrome P450 superfamily; CYP90B1" | | | |
| /allele="dwf4-1; dwf4-2; dwf4-3; dwf4-4" | /codon.start=1 | /codon.start=1 | | | |
| 990. 1055 | /product="steroid 22-alpha-hydroxylase" | /product="steroid 22-alpha-hydroxylase" | | | |
| /note="dwarf4" | /protein_id="PAC05093.1" | /protein_id="PAC05093.1" | | | |
| /allele="dwf4-1; dwf4-2; dwf4-3; dwf4-4" | /db_xref="GI:2935342" | /db_xref="GI:2935342" | | | |
| 990. 1055 | /translation="MFEFENHTPLPLLLPSLILFLILLKRRNRKTFNLPKGS | /translation="MFEFENHTPLPLLLPSLILFLILLKRRNRKTFNLPKGS | | | |
| /gene="DMF4" | GWPIRSETIGLYKPYTATLTGDFMOOHVSXGKTYRSNLEGPYIVSADAGLRPILO | GWPIRSETIGLYKPYTATLTGDFMOOHVSXGKTYRSNLEGPYIVSADAGLRPILO | | | |
| /note="dwarf4" | NEGLPECSYPRSIGIITGLKRWMLYUIMDKHNSRMRSLNPLSHARLFTLLIDVERH | NEGLPECSYPRSIGIITGLKRWMLYUIMDKHNSRMRSLNPLSHARLFTLLIDVERH | | | |
| /allele="dwf4-1; dwf4-2; dwf4-3; dwf4-4" | TLRYLDSMOONSIFSRADDEKKFTFNMAKHNSMDGEETFDLKEVYTFPKGVYS | TLRYLDSMOONSIFSRADDEKKFTFNMAKHNSMDGEETFDLKEVYTFPKGVYS | | | |
| 990. 1055 | APLNLPGTAYHKALQSRAITLKEFLERMERKHLIDIEDEDEEVKTEIDAESKSOH | APLNLPGTAYHKALQSRAITLKEFLERMERKHLIDIEDEDEEVKTEIDAESKSOH | | | |
| /gene="DMF4" | VRKRRDIDLGAWLKLSNLTSTEOIDLILISLFAGETSSVAIALAIFLQACPKAV | VRKRRDIDLGAWLKLSNLTSTEOIDLILISLFAGETSSVAIALAIFLQACPKAV | | | |
| /note="dwarf4" | ELREELHELITARAKELGESELMDWDVKKMDPQCVINETLRIGNVVFLHRAALQD | ELREELHELITARAKELGESELMDWDVKKMDPQCVINETLRIGNVVFLHRAALQD | | | |
| /allele="dwf4-1; dwf4-2; dwf4-3; dwf4-4" | RYGSDYIPSGMKATLPVTSVAVHLDNSRDOPNLNPNRMOONNGASSGSGSPSTGNC | RYGSDYIPSGMKATLPVTSVAVHLDNSRDOPNLNPNRMOONNGASSGSGSPSTGNC | | | |
| 990. 1055 | NYMFGGGRPLCACSELAKLEMAVFTIHLVLKRWLELAEDDQEPFAFPVDFPNGLPR | NYMFGGGRPLCACSELAKLEMAVFTIHLVLKRWLELAEDDQEPFAFPVDFPNGLPR | | | |
| /gene="DMF4" | VSRL" | VSRL" | | | |

REFERENCE
AUTHORS
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT
FEATURES
SOURCE

Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;
Brassicales; Brassicaceae; Arabidopsis.
1 (bases 1 to 84196)
Bloembergen, H., Mewes, H.W., Lemcke, K., Mayer, K.F.X., Quetier, F. and
Salanoubat, M.
Unpublished
2 (bases 1 to 84196)
EU Arabidopsis sequencing project.
Direct Submission
Submitted (25-JAN-2000) MIPS, at the Max-Planck-Institut fuer
Biochemie, Am Klopferspitz 18a, D-82152 Martinsried, FRG, E-mail:
lemcke@mips.biochem.mpg.de, mayer@mips.biochem.mpg.de, project
Coordinator: Marcel Salanoubat and Francis Quetier, Groupement
d'Interet Public, Centre National de Sequencage - GENOSCOPE, 2 rue
Gaston Creneau, BP191, 91006 Evry Cedex, France;
http://www.genoscope.cns.fr
On Jan 27, 2000 this sequence version replaced gi:6434247.
Information on performance of analysis and a more detailed
annotation of this entry and other sequences of chromosomes 3, 4
and 5 can be viewed at: http://www.mips.biochem.mpg.de/proj/thal/
location/Qualifiers
1. 84196
/organism="Arabidopsis thaliana"
/variety="Columbia"
/db_xref="taxon:3702"
/chromosome="3"
2983. 3338
/note="355 bp LTR"
3339. 7564
/note="LTR-transposon"
7565. 7922
/note="367bp LTR"
12339. 13197
/gene="T3A5.10"
join(12339..12665,12742..12803,12888..13055,13125..13197)
/gene="T3A5.10"
/note="Similarity to cyclin-dependent kinase inhibitor
(ICK1), Arabidopsis thaliana, EMBL:ATU9472_1"
/codon_start=1
/product="putative protein"
/protein_id="CAB62432.1"
/db_xref="GI:561966"
/translation="MAAVRRRRDVEENGVTYTVKRRKKEEVDLVESTRILSPCY
QATNRGGIVANSAGASSETSVIVRRDSPVEOCIEEDSSVCCSTSEKSKRR
IEFDLENNDDRETESMTYDLNKSSEMMNDSVAVEDVESRRRLKSLHEVY
KEALDEDFOVAEDLNKLKLECSMKYNFDEKDEPLOGGKRYEMVKLNP"
12339..12665
/gene="T3A5.10"
/number=1
12666..12741
/gene="T3A5.10"
/number=1
12742..12803
/gene="T3A5.10"
/number=1
12804..12887
/gene="T3A5.10"
/number=2
12888..13055
/gene="T3A5.10"
/number=3
13056..13124
/gene="T3A5.10"
/number=3
13125..13197
/gene="T3A5.10"
/number=4
complement(15858..15901)
15858..16470
/gene="T3A5.20"
complement(join(15858..15901,16014..16470))
/note="similarity to predicted proteins, Arabidopsis

thaliana"
/codon_start=1
/product="putative protein"
/protein_id="CAB62433.1"
/db_xref="GI:561967"
/translation="MARLEHNSMNGARISFNSPEVFEIRSEKSNMNNISNFS
MPSADFAFVYDTSMTIPADEFLGLKILPEKETSIVHRTGLGEELTEEGSMVGNF
SLRPLILSSSFSTKGTWRELLGLKRTIHVRSKTKDKVNEEVLSDHKIISGNVTRC
OAVDTR"
complement(15902..16013)
/number=1
complement(16014..16470)
/number=2
18303..19931
/gene="T3A5.30"
complement(18303..19931)
/number=1
complement(18303..19931)
/codon_start=1
/product="scarecrow-like 7 (SCL7)"
/protein_id="CAB62434.1"
/db_xref="GI:561968"
/translation="MAYTCIDSGNLMIAIQOLIKOKQOQSQHQOEEQDEPNPWN
PSFGFTLPSCGFSDPFQVNDPGPHFLEHHQNAVAASEEFDSEWMSLINDGAS
QTNDDPFIYGHDPFSPFSLAPSVLNVRKQDSASQOLPPPSAPAIWSPSPSQ
HPPPPPOPDNLNPIFKATHDYARKPEPTDPLIRIKESVSGDPIORGVYFAE
ALSHKTEPSSSSSSLEDFILSYKTNACCPYSKFAHLTANQALIEATNOSNIHI
VDFCITFOGLOMIALQALATRSCKPRTIRISGIPASLDGSPPSILATGNRLRDR
AAILDNEFYFVLTPIQLNGSSFRVDPDEVLVNFMELTKLIDETATVGLRLA
RLSLNRIYVTLGEVSLNRYEAFNRKNSLRFSAVESLEPNLDRSKERLYERYL
FGRRIMDLVRSDDDNKRGTFGLMEKEEQRVMEKAGFEPPVPRSNVAASOAKLLIM
NYSSTLYSLVSESEPGFISLANNVPLTVSSWR"
26093..28999
/gene="T3A5.40"
complement(join(26093..26338,27091..27200,27308..27386,
27478..27570,27723..28037,28136..28288,28374..28698,
28779..28999))
/note="Contains cytochrome P450 cysteine heme-iron ligand
signature AA455-464: prokaryotic membrane lipoprotein lipid
attachment site AA320-330"
/codon_start=1
/product="steroid 22-alpha-hydroxylase (DWF4)"
/protein_id="CAB62435.1"
/db_xref="GI:561969"
/translation="MPEFTEHTLPLILPELSLILFLILKRRNRKTRPNLPGRS
GMPLIGETIGLAKYTRATLTDPMQOHSKYGKIRYSLREPTIVSDAGLNPILO
NEGRLEFCSTYRSIGTIGKWSMLVGDMDRMSISLNTLSARKETILKDYERS
TLFVLDWMQNSIFSAQDEARKFTFNMAKIMSDPEEETEDLKREYVFMGVVS
APLNLPGTAYHKAQLQSRAATILKFTERKKEERLLIKEDQEEVEVTEDEAKMSOH
VRKORTDDLLGWLKHSNLSTEOILDLITSLFAGHETSVAALAIFFLOACPKAY
ELREBEHLERAKKELGESELMNDQYKKMFTQCIVINETRLGNVAFHLKRAKLOV
RYKGYDIPSGKXVLPVISAHLNDSRQDQPLFPMRMOONNKAASSGSGSPSTMN
NYMPFGGPRILCAGSELAKLEMAVFIHHLVLKFWELAEEDKPAFPFVDPNGLPIR
VSRIL"
complement(26093..26338)
/number=1
complement(26339..27090)
/number=1
complement(27091..27200)
/number=2
complement(27201..27307)
/number=2
complement(27308..27386)
/number=3
complement(27387..27477)
complement(27478..27570)
complement(27571..27722)
/number=4
complement(27723..28037)
/number=5
complement(28038..28135)

```

exon      /number=5
           /complement(28136..28288)
intron    /number=6
           /complement(28289..28373)
exon      /number=6
           /complement(28374..28698)
intron    /number=7
           /complement(28699..28778)
exon      /number=7
           /complement(28779..28999)
gene      /number=8
           /38307..41323
           /gene="T3A5.50"
exon      /complement(38307..38796)
           /number=1
           /complement(join(38307..38796,38877..39090,39189..39270,
40180..40242,40337..40401,40489..40564,40655..40734,
40805..40919,41225..41323))
           /note="Contains Eukaryotic putative RNA-binding region
RNP-1 signature AA179-186"
           /codon.start=1
           /product="rli snRNP 70K protein"
           /protein_id="CAB62436.1"
           /db_xref="GI:6561970"
           /translation="MGDSGDFLRNPNAAVQARAKVONRANVLIQLMGOSHPYGLTN
NLKLFPPRPPELEKPPPEPKCPYGMQFVNFAPGDPPEYAPKPEVELPSOKR
ERIKRLKLEGEKAEADLKDYNDNDPNATGDPEYKTLFVSLNVESSSEKIKREPS
YGPRIKRVHLYTDLTNPKGAYLEVMHTDMKAAKQADGOKIDGRVLYDVRGRT
VPMNRPRLGGGLGTSRYVGGEELVGEQOQGRISQSEPSRPREKRESEKERE
RSRLSHQPPERSRDRPREDKHNRDQGRDSDRSDRSDRSDRSDRSDRSDRSDR
GRDTSRDHNRDRKRKREYEGGEYHEGGGRSDRSDRSDRSDRSDRSDRSDRSDR
TDRYSRDYDMEEDDFRYEREYKRSKRSSEYVR"
intron    /number=1
           /complement(38797..38876)
exon      /complement(38877..39090)
intron    /number=2
           /complement(39091..39188)
exon      /number=2
           /complement(39189..39270)
intron    /number=3
           /complement(39271..40179)
exon      /number=3
           /complement(40180..40242)
intron    /number=4
           /complement(40243..40336)
Query Match      99.6%; Score 356.4; DB 14; Length 84196;
Best Local Similarity 99.7%; Pred. No. 3,8e-49;
Matches 357; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```

```

Db 25792 AACTTTACCTGTTTTGATCGAAGCAATACAAATTGTCAATACCAACACACAG 25735
RESULT 3
CNS073CX 1147 bp DNA STS 11-JAN-2001
LOCUS     clone BA0AB018F06 of library BA0AB from strain CLIB 210 of
DEFINITION Kluveromyces lactis, sequence tagged site.
ACCESSION AL427351
VERSION   AL427351.1 GI:12210545
KEYWORDS  STS.
SOURCE    Kluveromyces lactis.
ORGANISM  Kluveromyces lactis.
REFERENCE 1 (bases 1 to 1147)
           Bolotin-Fukuhara,M., Toffano-Nioche,C., Artiguenave,F.,
           Duchateau-Nguyen,G., Lemaire,M., Marmelisse,R., Montrocher,R.,
           Robert,C., Termier,M., Winkler,P. and Wesolowski-Louvel,M.
           Genomic Exploration of the Hemiascomycetous Yeasts: 11.
           Kluveromyces lactis
           FEBS Lett. 487 (1), 66-70 (2000)
TITLE     Kluveromyces lactis
JOURNAL   PUBMED
PUBMED   11152886
REFERENCE 2 (bases 1 to 1147)
           Souciet,J.L., Aigle,M., Artiguenave,F., Blandin,G.,
           Bolotin-Fukuhara,M., Bon,E., Brottier,P., Casaregola,S.,
           de-Montigny,J., Dujon,B., Durrens,P., Lepingle,A., Llorente,B.,
           Malpertuy,A., Neuvéglise,C., Ozler-Kalogeropoulos,O., Potier,S.,
           Saurin,W., Tekala,F., Toffano-Nioche,C., Wesolowski-Louvel,M.,
           Winkler,P. and Weissenbach,J.
           Genomic Exploration of the Hemiascomycetous Yeasts: 1. A set of
           yeast species for molecular evolution studies(1)
           FEBS Lett. 487 (1), 3-12 (2000)
TITLE     Kluveromyces lactis
JOURNAL   PUBMED
PUBMED   11152876
REFERENCE 3 (bases 1 to 1147)
           Genoscope.
           Direct Submission
           Submitted (08-SEP-2000) Genoscope - Centre National de Sequencage,
           2 rue Gaston Crémieux, CP 5706, 91057 EVRY cedex, FRANCE. (E-mail :
           segre@genoscope.cns.fr - Web : www.genoscope.cns.fr)
COMMENT   This STS is part of a random genomic sequencing program of thirteen
           yeast species: Saccharomyces bayanus var. uvarum, Saccharomyces
           exiguus, Saccharomyces servazii, Zygosaccharomyces rouxii,
           Saccharomyces kluyveri, Kluveromyces thermotolerans, Kluveromyces
           lactis var. lactis, Kluveromyces marxianus var. marxianus, Pichia
           angusta, Debaryomyces hansenii var. hansenii, Pichia sorbitophila,
           Candida tropicalis and Yarrowia lipolytica. Genomic inserts of 3 to
           5 kb were prepared and both extremities were sequenced. See
           keywords for description of this sequence and for the sequence of
           the other extremity of this insert.
FEATURES
  source
    1..1147
    /organism="Kluveromyces lactis"
    /strain="CLIB 210"
    /variety="lactis"
    /db_xref="taxon:28985"
    /clone="BA0AB018F06"
    /clone_1bp="BA0AB"
    1..1147
    /note="part of mitochondrial DNA"
    /evidence=not-experimental
  BASE COUNT 284 a 120 c 175 g 361 t 207 others
  ORIGIN
    Query Match      15.8%; Score 56.6; DB 53; Length 1147;
    Best Local Similarity 39.6%; Pred. NO. 1.6;
    Matches 90; Conservative 39; Mismatches 98; Indels 0; Gaps 0;

```

| | | | |
|--|----------|---|-------------|
| OY | 70 | cccaatgcatataaaatttggcataaatatcatatcattgcatttcgttttagctcg | 129 |
| Dd | 745 | TTTTTTT TTTT TTTATATATWAWMAAAAANTWAAWAAAGATAATTTTTTTTKTKGGGR | 804 |
| OY | 130 | gttcgagaaaagggttcgacttcgtaagaagtgcacgatcatatagatggagctagt | 189 |
| Dd | 805 | RATAAAAAAAMWAAGMAGKGTDMWMAAAATAGRTTMMWWMTAATTAATKKMWDTGCGGT | 864 |
| OY | 190 | tgaagtcttgacattgctatgatgttgttgcatttatagtgtcga | 236 |
| Dd | 865 | TTTTTTTTTTTAAKTTRKGSGGRRTTTTTKKKRKGSGKKKKWAGA | 911 |
| RESULT 4 | | | |
| AC004157 | LOCUS | AC004157 169546 bp DNA HTG | 12-AUG-2000 |
| DEFINITION | | Plasmodium falciparum chromosome 12 clone 3D7, *** SEQUENCING IN PROGRESS ***, 2 unordered pieces. | |
| AC004157 | AC004157 | AC004157 8 GI:9797712 | |
| KEYWORDS | SOURCE | HTG; HTGS-PHASEI. | |
| ORGANISM | | malaria parasite P. falciparum. | |
| REFERENCE | | Plasmodium falciparum | |
| AUTHORS | TITLE | Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium. | |
| JOURNAL | | 1 (bases 1 to 169546) | |
| REFERENCE | | Hyman,R.W., Fung,E.L., Qin,F., Rowley,D., Mao,J., Tanaki,T., Kundi,O.B., Conway,A.B. and Davis,R.W. | |
| AUTHORS | TITLE | Plasmodium falciparum 3D7 chromosome 12 unpublished | |
| JOURNAL | | 2 (bases 1 to 169546) | |
| | | Hyman,R.W., Qin,F., Fung,E.L., Conway,A.B. and Davis,R.W. | |
| | | Direct Submission | |
| | | Submitted (19-FEB-1998) Stanford DNA Sequencing and Technology Center, Stanford University, 855 California Avenue, Palo Alto, CA 94304, USA | |
| COMMENT | | On Aug 12, 2000 this sequence version replaced gi:8810447. | |
| | | * NOTE: This is a 'working draft' sequence. It currently | |
| | | * consists of 2 contigs. The true order of the pieces | |
| | | * is not known and their order in this sequence record is | |
| | | * arbitrary. Gaps between the contigs are represented as | |
| | | * runs of N, but the exact sizes of the gaps are unknown. | |
| | | * This record will be updated with the finished sequence | |
| | | * as soon as it is available and the accession number will | |
| | | * be preserved. | |
| | | * 1 23466: contig of 23466 bp in length | |
| | | * * 23467 23666: gap of unknown length | |
| | | * * 23667 169546: contig of 145880 bp in length. | |
| FEATURES | SOURCE | Location/Qualifiers | |
| | | 1..169546 | |
| | | /organism="Plasmodium falciparum" | |
| | | /db_xref="taxon:5833" | |
| | | /chromosome="12" | |
| | | /clone="PFYAC293" | |
| | | /clone="3D7" | |
| BASE COUNT | | 69871 a 15381 c 15705 g 68389 t 200 others | |
| ORIGIN | | | |
| Query Match | | 15.5% Score 55.6 DB 60 Length 169546; | |
| Best Local Similarity | | 51.6% Pred. NO.1.5 Indels 0 Gaps 0; | |
| Matches 127 Conservative 0 Mismatches 119 Indels 0 Gaps 0; | | | |
| OY | 1 | aaaaaaaaaagatgaaagatattctatctctctcttttttttgataatttaaca | 60 |
| Dd | 136495 | AAAAAAAAAAAAAAAAAGATGAACATTGAACTTTTGTGATGTTTAACTTTT | 136554 |
| OY | 61 | ttttttgccaatgatcatataaaaaatttgcataaatatcatatcattgcatt | 120 |
| Dd | 136555 | TTTTTTTTTCCTTTTGCCACACAAGAAATAATCATATTTCAATTTTGGAATTTT | 136614 |
| OY | 121 | ttagtctgggttgcgaaaagggttcgacttcgaaagtgcagatgatatatagattg | 180 |
| Dd | 136615 | TGAATATGGAATTTAAAAAACCTATAGCTGCATCAATTAATACAAATATACTGCCCTTTGG | 136674 |

Query Match 15.4% Score 55; DB 53; Length 1007;
Best Local Similarity 39.9%; Pred. NO. 3;
Matches 109; Conservative 37; Mismatches 127; Indels 0; Gaps 0;

DY 181 gagcaggttgacgcttggaccattgcattgatggatcgctgtgatattatagcgtcacact 240
DY 136735 GTAAATGAAAAAATTATTCACATTATTATAGACATTATATATTATTTATTTATATATA 136734

Oy 241 attaa 246
|||
Db 136735 TATATA 136740

RESULT 5
CNS06X9S/C CNS06X9S DNA STS 11-JAN-2001
LOCUS T3 end of clone AX0AA039F08 of library AX0AA from strain CBS 7064
DEFINITION of Pichia farinosa, sequence tagged site.
ACCESSION ALA19462
VERSION ALA19462.1 GI:12202640
KEYWORDS STS.
SOURCE Pichia farinosa.
ORGANISM Pichia farinosa
Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
Saccharomycetales; Saccharomycetaceae; Pichia.
REFERENCE 1 (bases 1 to 1007)
AUTHORS de Montigny,J., Spehner,C., Souclet,J., Tekala,F., Dujon,B.,
Wincker,P., Artiguenave,F. and Potier,S.
Genomic Exploration of the Hemiascomycetous Yeasts: 15. Pichia
sorbitophila
FEBS Lett. 487 (1), 87-90 (2000)
JOURNAL 11152890
PUBMED 2 (bases 1 to 1007)
REFERENCE Souclet,J.L., Aigle,M., Artiguenave,F., Blandin,G.,
Bojotin-Fukuhara,M., Bon,E., Brotier,P., Casaregola,S.,
de Montigny,J., Dujon,B., Durrens,P., Lepingle,A., Morente,B.,
Malpertuy,A., Neveuglise,C., Ozier-Kalogeropoulos,O., Potier,S.,
Saurin,M., Tekala,F., Toffano-Nioche,C., Wesolowski-Louvel,M.,
Wincker,P. and Weissendach,J.
Genomic Exploration of the Hemiascomycetous Yeasts: 1. A set of
yeast species for molecular evolution studies(1)
YEAST Lett. 487 (1), 3-12 (2000)
JOURNAL 11152876
PUBMED 3 (bases 1 to 1007)
REFERENCE Genoscope.
DIRECT SUBMISSION Submitted (08-SEP-2000) Genoscope - Centre National de Sequencage,
2 rue Gaston Cremlieux, CP 5706, 91057 EVRY cedex, FRANCE. (E-mail :
sequefgenosco.cns.fr - Web : www.genoscope.cns.fr)
COMMENT This STS IS part of a random genomic sequencing program of thirteen
yeast species: Saccharomyces bayanus var. uvarum, Saccharomyces
exiguus, Saccharomyces servazzii, Zygosaccharomyces rouxii,
Saccharomyces kluyveri, Kluyveromyces thermotolerans, Kluyveromyces
lactic var. lactic, Kluyveromyces marxianus var. marxianus, Pichia
angusta, Debaryomyces hansenii var. hansenii, Pichia sorbitophila,
Candida tropicalis and Yarrowia lipolytica. Genomic inserts of 3 to
5 kb were prepared and both extremities were sequenced. See
keywords for description of this sequence and for the sequence of
the other extremity of this insert.

FEATURES
source location/Qualifiers
1..1007
/organism="Pichia farinosa"
/strain="CBS 7064"
/db_xref="taxon:4920"
/clone="AX0AA039F08"
/clone_lib="AX0AA"
/note="end : T3"

STS
BASE COUNT 533 a 86 c 126 g 161 t 101 others
ORIGIN

BASE COUNT 335 a 117 c 96 g 498 t 121 others
ORIGIN

Query Match 15.0%; Score 53.6; DB 53; Length 1167;
Best Local Similarity 46.0%; Pred. No. 5;
Matches 122; Conservative 12; Mismatches 131; Indels 0; Gaps 0;

QY 1 aaaaaaaaaagatgaagattttatctctctctttttttttttgataatttaacta 60
|||||
Db 853 AAAAAAAAAAATATAAATAATATTTNNNTTTTNTTTTNTTTTNTTTTNTTWT 912
|||||
QY 61 ttctttgccaatgatataataaatttgataaataattattatgatactgttt 120
|||||
Db 913 TTTTWT 972
|||||
QY 121 ttgcttcgggttgagaaagggttcgacttcgaaagtgcagatgatattg 180
|||||
Db 973 TTTTTCGAAWMAAATAATATTTTATTTTNTTTTNTTTTNTTTTNTTTT 1032
|||||
QY 181 gagctaggttgactctttgacattgacatgtgtgtgattatagtcgcact 240
|||||
Db 1033 TAAATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTAAA 1092
|||||
QY 241 attaaccttaaatgggcttctat 265
|||||
Db 1093 ATTTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTAT 1117
|||||

RESULT 8
LOCUS CINS071LE 973 bp DNA SFS 11-JAN-2001
DEFINITION clone BA0AB003B01 of library BA0AB from strain CLIB 210 of
Kluyveromyces lactis, sequence tagged site.
ACCESSION AL425064
VERSION AL425064.1 GI:12208258
KEYWORDS STS.
SOURCE Kluyveromyces lactis.
ORGANISM Kluyveromyces lactis.
REFERENCE 1 (bases 1 to 973)
AUTHORS Bolotin-Fukuhara, M., Toffano-Nioche, C., Artiguenave, F.,
Duchateau-Nguyen, G., Lemaire, M., Marmesse, R., Montrocher, R.,
Robert, C., Terrier, M., Wincker, P. and Wesolowski-Louvel, M.
Genomic Exploration of the Hemiascomycetous Yeasts: II.
Kluyveromyces lactis
FEMS Lett. 487 (1), 66-70 (2000)

JOURNAL
PUBMED 11152886
REFERENCE 2 (bases 1 to 973)
AUTHORS Souciet, J.L., Aigle, M., Artiguenave, F., Blandin, G.,
Bolotin-Fukuhara, M., Bon, E., Brothier, P., Casaregola, S.,
de-Montigny, J., Dujon, B., Durrens, P., Lepingle, A., Llorente, B.,
Malpertuy, A., Neuvéglise, C., Ozier-Kalogeropoulos, O., Potier, S.,
Saurin, W., Tekala, F., Toffano-Nioche, C., Wesolowski-Louvel, M.,
Wincker, P. and Weissenhach, J.
Genomic Exploration of the Hemiascomycetous Yeasts: 1. A set of
yeast species for molecular evolution studies(1)
FEMS Lett. 487 (1), 3-12 (2000)
JOURNAL
PUBMED 11152876
REFERENCE 3 (bases 1 to 973)
AUTHORS Genoscope.
TITLE Direct Submission
SUBMITTED (08-SEP-2000) Genoscope - Centre National de Séquençage,
2 rue Gaston Crémieux, CP 5706, 91057 EVRY cedex, FRANCE. (E-mail :
seqref@genoscope.cns.fr - Web : www.genoscope.cns.fr)
This STS is part of a random genomic sequencing program of thirteen
yeast species: Saccharomyces bayanus var. uvarum, Saccharomyces
exiguus, Saccharomyces servazzii, Zygosaccharomyces rouxii,
Saccharomyces kluyveri, Kluyveromyces thermotolerans, Kluyveromyces
lactis var. lactis, Kluyveromyces marxianus var. marxianus, Pichia
angusta, Debaryomyces hansenii var. hansenii, Pichia sorbitophila,
Candida tropicalis and Yarrowia lipolytica. Genomic inserts of 3 to

5 kb were prepared and both extremities were sequenced. See
keywords for description of this sequence and for the sequence of
the other extremity of this insert.

FEATURES
Source
Location/Qualifiers
1. 973
/organism="Kluyveromyces lactis"
/strain="CLIB 210"
/variety="lactis"
/db_xref="taxon:28985"
/clone="BA0AB003B01"
/clone_lib="BA0AB"
1. 973

BASE COUNT 505 a 92 c 23 g 225 t 128 others
ORIGIN

Query Match 14.9%; Score 53.4; DB 53; Length 973;
Best Local Similarity 40.9%; Pred. No. 5.5;
Matches 121; Conservative 25; Mismatches 150; Indels 0; Gaps 0;

QY 1 aaaaaaaaaagatgaagattttatctctctctttttttttttgataatttaacta 60
|||||
Db 424 AAAAAAAAAAATATAAATAATATTTTAAATTTTATTTTAAATATAATTTTMAAA 365
|||||
QY 61 ttctttgccaatgatataataaatttgataaataattattatgatactgttt 120
|||||
Db 364 TTTTATTTATTTTGTATTTTWTATATATATAATTTTATTTTATTTTATTTTAT 305
|||||
QY 121 ttgcttcgggttgagaaagggttcgacttcgaaagtgcagatgatattg 180
|||||
Db 304 TAAADGAAATWTAAGDWTAAATTTTWTWTATTTGNNNNNNNGNNNTKGGCGTGGG 245
|||||
QY 181 gagctaggttgactctttgacattgacatgtgtgtgattatagtcgcact 240
|||||
Db 244 GGTBTYKAGATGTGTTTTTTTTTTTCTTTTTTTTTTTTTRBBSVSKKKTKKRTTB 185
|||||
QY 241 attaaccttaaatgggcttctataagcccaattattatgataatacaag 296
|||||
Db 184 BBTBTBTACTTTTNTTATTTNTTNGCVTAAGTATKATGATTAATTAATTAAG 129
|||||

RESULT 9
LOCUS CINS061TX 573 bp DNA SFS 10-JAN-2001
DEFINITION T7 end of clone XMT0AA001E06 of library XMT0AA from strain CBS 4311
of Saccharomyces servazzii, sequence tagged site.
ACCESSION AL404635
VERSION AL404635.1 GI:12166351
KEYWORDS STS.
SOURCE Saccharomyces servazzii.
ORGANISM Saccharomyces servazzii.
REFERENCE 1 (bases 1 to 573)
AUTHORS Casaregola, S., Lepingle, A., Bon, E., Neuvéglise, C., Nguyen, H.,
Artiguenave, F., Wincker, P. and Gaillardin, C.
Genomic Exploration of the Hemiascomycetous Yeasts: 7.
Saccharomyces servazzii
FEMS Lett. 487 (1), 47-51 (2000)

JOURNAL
PUBMED 11152882
REFERENCE 2 (bases 1 to 573)
AUTHORS Souciet, J.L., Aigle, M., Artiguenave, F., Blandin, G.,
Bolotin-Fukuhara, M., Bon, E., Brothier, P., Casaregola, S.,
de-Montigny, J., Dujon, B., Durrens, P., Lepingle, A., Llorente, B.,
Malpertuy, A., Neuvéglise, C., Ozier-Kalogeropoulos, O., Potier, S.,
Saurin, W., Tekala, F., Toffano-Nioche, C., Wesolowski-Louvel, M.,
Wincker, P. and Weissenhach, J.
Genomic Exploration of the Hemiascomycetous Yeasts: 1. A set of
yeast species for molecular evolution studies(1)
FEMS Lett. 487 (1), 3-12 (2000)
JOURNAL
PUBMED 11152876
REFERENCE 3 (bases 1 to 573)
AUTHORS Genoscope.

JOURNAL

Submitted (18-JUN-1998) Stanford DNA Sequencing and Technology Center, Stanford University, 855 California Avenue, Palo Alto, CA 94304, USA

COMMENT

On Aug 12, 2000 this sequence version replaced g1:8810455.

* NOTE: This is a 'working draft' sequence. It currently consists of 4 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown. * This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.

FEATURES

source

1. 187308: contig of 187308 bp in length
* 187309 187508: gap of unknown length
* 187509 257820: contig of 70312 bp in length
* 257821 258020: gap of unknown length
* 258021 307235: contig of 49235 bp in length
* 307236 307455: gap of unknown length
* 307456 310779: contig of 3324 bp in length.
Location/Qualifiers
1. 310779
/organism="Plasmodium falciparum"
/db_xref="taxon:5833"
/chromosome="12"
/clone="PFYACB8-420"
/clone="3D7"

BASE COUNT 126134 a 29409 c 30056 g 124579 t 601 others
ORIGIN

Query Match 14.7%; Score 52.8; DB 60; Length 310779;
Best Local Similarity 53.1%; Pred. No. 4.1;
Matches 136; Conservative 0; Mismatches 117; Indels 3; Gaps 1;

OY 1 aaaaaaaagatgaagatattttatctctcttttttttgataatttaacaa' 60
Db 7626 AAAATTTAATAACCAATTAACCTTATTTGTTCTCTTATTTAATAAATTA 7685
OY 61 tttttttgc---caatgatataaaatttgataataatttgatttgcgt 117
Db 7686 AATTTTGTGCAATTTAAAAAACAATAAATGTAATATTTTGAATATTTT 7745
OY 118 ttttaagtcgggtttggaaggggttcgaacttcgaagagtgacgtatagat 177
Db 7746 AATTATTTTAAAAAATAATATATATTCAGTTTCAATGAGAAATATATTAATA 7805
OY 178 tggagagcaggttcagcttcgtgacatttgatggatggttgattatagtgac 237
Db 7806 TTGGAGTTAATGTTGTATATTACAAATGTAATCATATTTTACATTAGTATTTAA 7865
OY 238 actaataacttaaa 253
Db 7866 TTTATTTAATAATTA 7881

RESULT 12

AC014610

LOCUS 24091 bp DNA 16-NOV-1999
DEFINITION Drosophila melanogaster, *** SEQUENCING IN PROGRESS ***, in ordered pieces.

AC014610.1 GI:6436725

ACCESSION AC014610.1
VERSION GI:6436725
KEYWORDS HTG: HTGS, PHASE2.
SOURCE fruit fly.

ORGANISM

Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.

1 (bases 1 to 24091)

Adams, M. and Venter, J.C.

REFERENCE
AUTHORS Direct Submission
TITLE Submitted (16-NOV-1999) Celera Genomics, 45 West Gude Drive,
Rockville, MD, USA

COMMENT

This sequence was identified as CDM:10214343 by the submitter.
For further information on this sequence e-mail to fly@celera.com.

* NOTE: This is a 'working draft' sequence.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.

FEATURES

source

1. 24091

/organism="Drosophila melanogaster"

/db_xref="taxon:7227"

BASE COUNT 7749 a 4270 c 4283 g 7789 t
ORIGIN

Query Match 14.7%; Score 52.6; DB 63; Length 24091;
Best Local Similarity 46.8%; Pred. No. 5.5;
Matches 166; Conservative 0; Mismatches 189; Indels 0; Gaps 0;

OY 1 aaaaaaaagatgaagatattttatctctcttttttttgataatttaacaa 60
Db 18191 AAAAGTCAGAAAGTCATTAATTAATTAATTAATTAATTAATTAATTAAT 18250
OY 61 ttttttgcacatgatataaaatttgataataatttgatattcgttt 120
Db 18251 ATATATATAGAAATATATATATTTTAAACAAATGATGCTTATATATCAGTTT 18310
OY 121 ttatgtcgggttggaaggggttcgaacttcgaaagtgagatgatattgg 180
Db 18311 GAAAAAATTTTAAAGTGTATTTATTTTGAAGAAATGATCTTAAATATTTGTT 18370
OY 181 gagcaggttcagcttcgtgacatttgatggatggttgattatagtgacact 240
Db 18371 AATTAATTTTTCACCTTGTGAATATAGATCGATTATATTAATAAAGTCATTTTTCG 18430
OY 241 attaaccttaaatgagcttcctataagggccaattatatacgaattataacaaatgac 300
Db 18431 GCCAGACCAAAAGTCCGTTTACGAATCGAAAAATTTTACGATTTGAAAAACGAAG 18490
OY 301 aactttacttcgttttttgatccgaagcaataacaaatgtaataacaaacac 355
Db 18491 CATGCAATTTTCTATTTTCAATGAATAATTTGAAATTCCTTACGAATTCGATTAC 18545

RESULT 13

AE002751/c

LOCUS 106993 bp DNA INV 06-OCT-2000
DEFINITION Drosophila melanogaster genomic scaffold 142000013386031, complete sequence.

AE002751

AE002751.2 GI:10729377

ACCESSION AE002751
VERSION GI:10729377
KEYWORDS HTG: HTGS, PHASE2.
SOURCE fruit fly.

ORGANISM

Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.

1 (bases 1 to 106993)

Adams, M.D., Celisner, S.E., Holt, R.A., Evans, C.A., Gocayne, J.D.,

Amanatides, P.G., Scherer, S.E., Li, P.W., Hoskins, R.A., Galie, R.F.,

George, R.A., Lewis, S.E., Richards, S., Ashburner, M., Henderson, S.N.,

Sutton, G.C., Wortman, J.R., Yandell, M.D., Zhang, Q., Chen, L.X.,

Brandon, R.C., Rogers, Y.H., Blazey, R.G., Champe, M., Pfeiffer, B.D.,

Wan, K.H., Doyle, C., Baxter, E.G., Helt, G., Nelson, C.R., Gabor

Miklos, G.L., Abril, J.F., Agbayani, A., An, H.J.,

Andrews, Pannkock, C., Baldwin, D., Balley, R.M., Basu, A.,

Baxendale, J., Bayraktaroglu, L., Beasley, E.M., Beeson, K.Y.,

Benos, P.V., Berman, B.P., Bhandari, D., Bolshakov, S., Borikova, D.,

Botchan, M.R., Bouck, J., Brokstein, P., Brothier, P., Burks, K.C.,

Busam, D.A., Butler, H., Cadieu, E., Center, A., Chandra, I.,

Cheriy, J.M., Crawley, S., Dahlke, C., Davenport, L.B., Davies, P.,

de Pablos, B., Delcher, A., Deng, Z., Mays, A.D., Dew, I., Dietz, S.M.,

Dodson, K., Doup, L.E., Downes, M., Dugan-Rocha, S., Dunkov, B.C.,

Dunn, P., Durbin, K.J., Evangelista, C.C., Ferraz, C., Fiertera, S.,

FILESCHMANN, W., FOSLER, C., GABRIELIAN, A.E., GARG, N.S.,
GELBART, W.M., GLASSER, K., GLODEK, A., GONG, F., GORRELL, J.H., GU, Z.,
GUAN, P., HARRIS, M., HARRIS, N.L., HARVEY, D., HELMAN, T.J.,
HERNANDEZ, J.R., HOUCK, J., HOSTIN, D., HOUSTON, K.A., HOWLAND, T.J.,
WEI, M.H., ILEGWAM, C., JALALI, M., KALUSH, F., KARPEN, G.H., KE, Z.,
KENNISON, J.A., KETCHUM, K.A., KIMMEL, B.E., KODIRA, C.D., KRAFT, C.,
KRAVITZ, S., KULP, D., LAI, Z., LASKO, P., DEL, Y., LEVITSKY, A.A.,
LI, J., LI, Z., LIANG, Y., LIN, X., LIU, X., MATTEL, B., MCINTOSH, T.C.,
MCLEOD, M.P., MCPHERSON, D., MERKULOV, G., MISHINA, N.V., MOHARRY, C.,
MORRIS, J., MOSHREFI, A., MOUNT, S.M., MOY, M., MURPHY, B., MURPHY, L.,
MUZNY, D.M., NELSON, D.L., NELSON, D.R., NELSON, K.A., NIXON, K.,
NUSKERN, D.R., PACLEB, J.M., PALAZZOLO, M., PITMAN, G.S., PAN, S.,
POLLARD, J., PULI, V., REESE, M.G., REINERT, K., REMINGTON, K.,
SUNDERS, R.D., SCHEELER, F., SHEN, H., SHUE, B.C., SIDEN-KIAMOS, I.,
SIMPSON, M., SKUPSKI, M.P., SMITH, T., SPIER, E., SPRADLING, A.C.,
STAPLETON, M., STRONG, R., SUN, E., SVIRSKAS, R., TECTOR, C., TURNER, R.,
VENTER, E., WANG, A.H., WANG, X., WANG, Z.Y., WASSERMAN, D.A.,
WEINSTECK, G.M., WEISSENBAUGH, J., WILLIAMS, S.M., WOODAGE, T.,
WORLEY, K.C., WU, D., YANG, S., YAO, Q.A., YE, J., YEN, R.F.,
ZAVERT, J.S., ZHAN, M., ZHANG, G., ZHAO, Q., ZHENG, L., ZHENG, X.H.,
ZHONG, F.N., ZHONG, W., ZHOU, X., ZHU, S., ZHU, X., SMITH, H.O.,
GIBBS, R.A., MYERS, E.W., RUBIN, G.M. and VENTER, J.C.
The genome sequence of *Drosophila melanogaster*
Science 287 (5461), 2185-2195 (2000)
20196006
2 (bases 1 to 106993)
ADAMS, M.D., CELINKER, S.E., GIBBS, R.A., RUBIN, G.M. and VENTER, C.J.
Direct Submission
Submitted (21-MAR-2000) Celera Genomics, 45 West Gude Drive,
Rockville, MD, USA
On Oct 9, 2000 this sequence version replaced gi:7289853.
Location/Qualifiers
1. 106993
/organism="Drosophila melanogaster"
/db_xref="taxon:7227"
/chromosome="2R"
/map="41C-41D"
/map="41F1-41F1"
join(<57498, 57614, 57686, .>57928)
/gene="CG17482"
/product="CT38651"
/db_xref="FLYBASE:FBan0017482"
/db_xref="FLYBASE:FBgn0032994"
<57498, .>57928
/gene="CG17482"
/db_xref="FLYBASE:FBan0017482"
/db_xref="FLYBASE:FBgn0032994"
join(57498, 57614, 57686, .57928)
/gene="CG17482"
/note="CG17482 gene product"
/codon_start=1
/db_xref="FLYBASE:FBan0017482"
/db_xref="FLYBASE:FBgn0032994"
/protein_id="AAFA5460.1"
/db_xref="GI:7289853"
/translation="MDAGSIIIVCYMSRTGSNEMVTGKTNETSEQYPELCTARSDND
ERNPDYEMEVMTFGASCSPSSAHNVKTLTLTKFMPDSDSRAVFTIDYHYVDYVSF
VTERSSPYEYELGPG"
join(68140, .68292, 72127, .72917, 79504, .80049, 81286, .82015,
87046, .87370)
/gene="p120ctn"
/product="CT38657"
/db_xref="FLYBASE:FBan0017484"
/db_xref="FLYBASE:FBgn0015587"
<68140, .>87370
/gene="p120ctn"
/note="CG17484"
/db_xref="FLYBASE:FBan0017484"
/db_xref="FLYBASE:FBgn0015587"
join(72160, .72917, 79504, .80049, 81286, .82015, 87046, .87351)
/gene="p120ctn gene product"
/codon_start=1

/db_xref="FLYBASE:FBan0017484"
/db_xref="FLYBASE:FBgn0015587"
/protein_id="AAFA5461.1"
/db_xref="GI:7289855"
/translation="MQSORFSGFQWYDQPSAAGYSSTPLTYITKPTGCAITKVAHPCSO
NKIETNSTDSNIDPIENHPLATVRYIYQAGYEDTSEYGAGLTCGIDSETYTAAVVP
YCYTSDANVYGIQNTTSSIKPFSGRPNDINNGEAVADSCQRTFKSAEKLPOFM
SSINTVPORLQEKDDYIEGSENDSCIPMRNDPMLISYIFLSNPSSAIFAMAAYLQ
HICVMDPNKORTSLGIPPLVLLSYDSPEIDIKNCGALRNLSYRONDENKRGIK
NAGCIALVHLRCSQETPEYKELVLYGLVMMSSCEDIKRSIIDELAAVCSYIKPKS
GMDVACGECFCSYVPFMAAGSVLVNYSAGEHACCLRNCEHVECLLYVYRTSIEN
NIGKIVKNCILRNLSYRQEVDDNYDHPITPERYIPSCNPEETLEYAAGALQ
NLSACYQSPIDIRATYKKEGKLPILVELLMEYDRAVCAVATLRNLAIIDQNRKL
GKXAMRDVLRKPSGNVQHDQNSDDITAVLATINVEIKNPEFSRSLDSCGIDL
MLNTRREKTESCVLKFAQVLYTMMQHNELRDYKKNKGEODFVSKHFTANTPPS
SPNNVNDLIRPMAOSQRTREDTETIGRSTIYSANDSGAVVMSNEMISEMKRA
EMQKVNDVLAIRLTYQYEDGILKRNKRLGKTIKAKVHARVDEKVGDEGSPKTLT
VSEPRKFEFGINTQSGLPVSGIKKPEEOIGRETRSGRRL"
complement(join(<75533, .75970, 76088, .76363))
/gene="CG17479"
/product="CT38649"
/db_xref="FLYBASE:FBan0017479"
/db_xref="FLYBASE:FBgn0032996"
complement(<75533, .>76363)
/gene="CG17479"
/db_xref="FLYBASE:FBan0017479"
/db_xref="FLYBASE:FBgn0032996"
complement(join(75692, .75970, 76088, .76363))
/gene="CG17479"
/note="CG17479 gene product"
/codon_start=1
/db_xref="FLYBASE:FBan0017479"
/db_xref="FLYBASE:FBgn0032996"
/protein_id="AAFA5464.1"
/db_xref="GI:7289858"
/translation="MASIAVRAVNTENSKYSVEVRAGEDTGAEGYRYTPGIRLENG
EGSSIRSFAGIKTVPPSPKGLSQSPAALSLCRLQWIKRROEHEHLEFTHAVRL
KTEVALSETYRVNPGCKWVSDQKGTISTAVAGARTISDDAAIVIAGLVFLEDEL
LKKDARRSSLDNMOMANDNSSKR"
91799, .>93503
/gene="CG17486"
/product="CT38665"
/db_xref="FLYBASE:FBan0017486"
/db_xref="FLYBASE:FBgn0032997"
<91799, .>93503
/gene="CG17486"
/db_xref="FLYBASE:FBan0017486"
/db_xref="FLYBASE:FBgn0032997"
91809, .93503
/gene="CG17486"
/note="CG17486 gene product"
/codon_start=1
/db_xref="FLYBASE:FBan0017486"
/db_xref="FLYBASE:FBgn0032997"
/protein_id="AAFA5462.1"
/db_xref="GI:7289856"
/translation="MCGICFCSYVNNVPLNSFNISALKVLTKNRGPDVQEDVADCF
GKILPAGFVLMOGSEYQKQVYEDDPIFLNGDIYNTSKPEYMSOTMTIAEPLABCR
CKEQNLIKILKREGECHLIIYKRBIDILFESDIALCRNSLLIERICNGQLSTSHF
EDKRISLEPPLGLFRVYKLDNSCVLYPQAPLNNSTOLLRNLDAIGKTAVESPM
SDPMWLNCPAFSYNFEYQIKNNVLYXNLIDQSEIKYTLISLHLTLSDSINKNR
NMAFQRLCMQKLNISPCRHAKCLIFSGGLCTTIALANQFVNPETIELINVF
ESVKGQNSIEKLPDVBDRKSLVSNVLEKQCEPCWNLKLVNTYRLQLQHLSTRK
HLIYLPDIYVLESLGCAFWASCHSTAVAILGSGADELFGGYPRNYSRQCN
DLEROLAVONELEERDMORISARNLARDRIATGTARSPFIEENRVKTRISLEYO
KCCGFPEVGGDKLLRLTYGQIGLDRVDVFLKRAIQFGRIRANKQNGSHSDNL"
complement(join(93644, .93777, 94377, .94539, 94544, .>95261))
/gene="CG17478"
/product="CT38647"
/db_xref="FLYBASE:FBan0017478"
/db_xref="FLYBASE:FBgn0032998"
complement(<93644, .>95261)
/gene="CG17478"

TITLE
 Direct Submission
 JOURNAL
 Submitted (24-SEP-1998) P.falciparum Genome Sequencing Consortium,
 The Sanger Centre, Wellcome Trust Genome Campus, Hinxton, Cambridge
 CB10 1SA, UK
 COMMENT
 On Apr 28, 2000 this sequence version replaced gi:5731895.
 For more information about this sequence or the Malaria Project,
 see <http://www.sanger.ac.uk/Projects/P-falciparum>.
 FEATURES
 source
 1. 66441
 /organism="Plasmodium falciparum"
 /strain="3D7"
 /db_xref="taxon:5833"
 /chromosome="1"
 1. 579
 /note="telomeric repeats"
 join(29733..34985,36111..37349)
 /gene="var"
 /note="MAL1P4.01"
 join(29733..34985,36111..37349)
 /gene="var"
 /note="MAL1P4.01, var gene, len: 2163 aa, similarity: eg
 to Q25733 PFEMP1 variant 1 of strain MC (2924 aa) fasta
 scores: opt: 3681, E(): 0, 35.8% identity in 3011 aa
 overlap"
 /codon_start=1
 /product="var, MAL1P4.01"
 /protein_id="CAB89209.1"
 /db_xref="GI:7670005"
 /translation="MTTSSGGGAAGSSGEEDAHVLDDEGOVYNEKERYANSKIT
 YKALKGDSQASLSELAGYPCALREYKHTNGGGRKRYPCETELGEVEYRFS
 TLGGQCNKKLEGKNIYKAGDYACAPRRLHLCSHLESIOYNNYNSGNKHLVD
 VCAAAKEGDSIKNYPYRTYPTDINSQCLTVARSEADIGDYGKDIKLGNDOS
 TQRIILENNKIDIFAKIHSYDSTSGSNGALORRYKQTDNYIELREDMAYLNDOW
 KATCNMGGRYFROTGSGEMAKDKCRCDKVPYTFDVPYPLRFEEMADFCR
 LRKKRLDKAKKCGDSNDYCDLNRDCTQITRGNHEVEKDCGCQSCAHFVN
 WIDQKLEFEQEKRYEIKRKHPTTIIITANKRTTINLVYEFKQIQEYGDV
 ENFLQKNEQIICNOPYNDESSIDINFKSJKDIDISHTYCOACPGCGARRGKQ
 EPERKCGTKTYDPRKTTNIPILPTYSQSIKRYKFCNGNGNAPATPARGGQ
 IKMNOCHYEGDNNDNCEVEGWEKFEKGNVSNAPFMKWHMLDMSMOGRHNGC
 INKNDMNTCKNSCKRCECFKRWYDOKKNMEWAIKDHFKKONIATFECODRGLQK
 VLLIDPLAKDESTPEKENKYSAEKAEIKHLQMLQOAGVDDPAFARPCTEDGAEOD
 TTMOKLRENDTECKCDKPPAPAGRGADPHDVOOPRPGSGPTDAN
 DEDDDDDDDDEDEGEAKEEDEEOPAEHGEAKKEPOKEVAPRPKDYNVC
 NIVANNVPTDSSLOAACSILKYGNAPTSKCVPTSGTSDTVSGDTGSGICVPR
 RRLVYPLTFLTGSDSTOASQAEVOTQARGNTDKPSGSEAAOGDGVSKPOKA
 LLKAFESAAYETFLDMRYKKIKERKKKKKTYEIOIYESTDYDEEKKPOEELKQ
 IIPDEFKROMFYTLGDKDILYSGDYNGGNEKIKKALINYPKIRBOSSDNRPB
 RSVATPSTSDKDPOTMNAHAPSTIMNAVCAIYDINSGGEGKTTTITODNLKTALM
 DENGRKPLKTYOYDSVTIGASGAKPQTKAPGTGDPPLQFVLRLPYFRYLEMGON
 FCKRTEMLEKIKYEKVGQGRGGRKQTPQSCYGRNCDQLDNDSTDLKPCG
 GRECRKRYKMEKEKEFTKOSNYEOKTKCOSEKSAKGNINNEFCGOGCDTA
 GDFLNLKSGPCKKENGKNOEDEINFKDEKTFGHENYCAPCPVFFDICKKDCRRA
 SNMNCNGKDTTADIKIMSSSEVNMALVSDNTNKFSGDLCKCAHAFKGIKENK
 WSCNVCYGVACRKKKNGEKGSENNQIITIGLVTHVONLDDYKIRTKLPC
 RNNDESVKCIKDCVKKWKKTEEMPIRDRYLEPYSVSDGVNKSLSRSMELIPL
 MDLNGEAKIOELNKLRSYECNADNSQKGPRIKIVELLELEKDKANKCTQS
 GTDCHSTPLEDDEPLETEBENTEOINCPTEPOPEKDEGGEAAPTAEESTPA
 TSEGTENOSPPPPAP
 PALMSSTIMMSIGIGAPATYELIKKTKKTSVCLPOLDIPONDYCIPTLKSNRI
 PYRSGYKAKTYIYMEGSSGDEKATAMSDTDTLTSSESEIEELINDIYVAPARK
 TLEIVLEPSGNNTASGKNTPSDTONDIPSDTPTPTDNEWNTLADFTSOYLOSE
 QPKVPNDYKSGDIPLNTOPNTLYFNKPEEPFTSIHDRLYLGEQISINIHSTNT
 MDPKRYVSNVSGIDLINDALNGDYIDIELIKRKENELEGTNHWKOTSHSVAKL
 NSDPIHNOLELHKLDRHNMCKKNDNRSLAKLEEMENETHSGTNPSPSKRL
 NTQYSIOIDMDHEKMEFTNMDTILDDLTNYENPYDYVDDIYDVNDHDASTVDSN
 MDPVRSVQIEMEDVNTKLVEKFTIADVMDI"
 34985..34991
 /note="potential splice donor sequence, aag/gtaata"
 36101..36110
 /note="potential splice acceptor sequence"
 complement(join(39205..40146,40377..40430))
 /gene="r1fin"

CDS
 /note="MAL1P4.02"
 complement(join(39205..40146,40377..40430))
 /gene="r1fin"
 /note="MAL1P4.02, r1fin, len: 331 aa, similarity: most to
 097325 PF0010C protein (326 aa), fasta scores: opt: 1526,
 E(): 0, 72.2% identity in 331 aa overlap"
 /codon_start=1
 /product="r1fin, MAL1P4.02"
 /protein_id="CAB89210.1"
 /db_xref="GI:7670006"
 /translation="MKIRYINILFELPLNLIYNQNHKSTPTPTNHTOTRLICE
 CELYSPANNDDAEKRVQOPEDTQTFRHEHYDERMKTQKCEQCDKRIQIKILK
 DKLEKLEMDKATLQTDIOSDIPICLSEKSELYEKVCGCLRCAGVGLGAPWSLY
 SGLGAAVNTVYVOTALQKGLGAGVAKGIEGLRPSGGLIPISVIONLNNHTYDI
 AKYITFEKSVNSTRCAVKEHSPSTYSINEMNLKSKRAGIAEAAAMAKITTEGVL
 DAPIRGUTYSNATASVAIVYLVWIIITILIRRRKKMKKLYILLEE"
 complement(40147..40156)
 /note="potential splice acceptor sequence"
 complement(40371..40376)
 /note="potential splice donor sequence, ttg/gtaatt"
 complement(join(42589..43839,43997..46729))
 /gene="MAL1P4.03"
 complement(join(42589..43839,43997..46729))
 /gene="MAL1P4.03"
 /note="MAL1P4.03, var-like hypothetical protein, len: 1327
 aa, similarity: most to 96294 PFEMP1 fragment (431 aa),
 fasta scores: opt: 1852, E(): 0, 67.0% identity in 440 aa
 overlap"
 /codon_start=1
 /product="var-like protein, MAL1P4.03"
 /protein_id="CAB89211.1"
 /db_xref="GI:7670007"
 /translation="MALKKGINSKLSARVULEINAEIKVKNREKRYNDKLGSL
 MEARPSGLSSPDDIRSGYDSSLDHKEFTNTNNGYPPARNPCDNGNRRPSDGE
 SKGSDIKRGNENSNAGACAPFRKRNQDNLETLINKNTENHDLGNLVAKTE
 GESTLVNHPNDIDGNKSSICTSLARSPADIGDILRGDMKRRKHDIENGLREK
 KIKYGLKNNGAREHYKEVNGNRYIKLRDMWTARNDYWKAITCEAPENAYITRRID
 GGDLENLITPFGCHDPPYVYIDPRLKFMSEMSRYFCNVILNKETIDENNCKOC
 ESMRCHNDTGEKCKKQECQIFELKESYKNOPOKSKWYELVYKASTNITON
 SSAPERYRRRRRGDPPDNDQILFKYIENNEKQESIGKYLKDSKSHGNENY
 DNTGSGNSRNASFTTPEKFRACCKIPNLEKCPNEKNVCTRFDFYSCYTSIFK
 NDLESMNNSGVNKNENDNGVLVPPRRNLCTINLSKVDYKMKENDPEKEDULNAFS
 QGKILGRKYSNYSNAYEAMKFSADYSDIVKGDMMNDLKKLKELTMLTEKGD
 ISVBRKTMWDMKNVYVNMALCGKTEENQOLNMSMCNDIDQFLMLTEMAQ
 OYCEKILIKAHIIITKCDIYEGGRHSMVDITVECKRLFDIYEMFRYRNOMKGL
 SEYKIKIKKAIINVTNIPSEBACASVYTKHNGCICNRMEDYHKNINNOEMLKEM
 INIIRKPTDQRIQDNLNNSMENINTSVKTAVDTTDIYSYGLAGTNGVATIGLOAG
 DFLCKRIQDILNEFMKPEYKGLDYSKRNLTIEDPNTMVPAGIGVALTLGLLKRMT
 KPKHEVMIRILLOMSQNKYGIPTTPKSPKTYIPVASORKCTTYIYMEGSGDEKYG
 NISSSDITSSSEYEEMDINDIYTSRKYTLLEVLEPSKSGNTGPSKGDYVGD
 DMVTTNTFTDEESEELKHDFISMLQNOPDVNDYKSGDIPENOTPNLYFPNNOE
 KPTTISIDHRLYGESEISYNIHSTSMOPKRVSNVNVYSGLDILINDTSCGHIT
 YDEVLKRENELEFETNPKRNTSNNSVAKLNTSDPINNOGLDILHKLDRHRMCKRNT
 KEDILNKLNDQMNKDNVNGDIPNDNKLNTDVSIIQDIDENKCKKDLSDNTLLDDM
 EDIYLYVNDNDNDNDPSVYDISMDHNKVDVDPKRVHIEMLKILNNTSNGSLQRP
 ISDWNTI"
 complement(43840..43849)
 /note="potential splice acceptor sequence"
 complement(43991..43996)
 /note="potential splice donor sequence, aag/gtaatt"
 join(50585..50638,50794..51858)
 /gene="MAL1P4.04"
 /note="MAL1P4.04"
 join(50585..50638,50794..51858)
 /gene="r1fin"
 /note="MAL1P4.04, r1fin, len: 372 aa, similarity: to
 r1fins, most to 097327 PF0035W protein (374 aa), fasta
 scores: opt: 1216, E(): 0, 55.9% identity in 374 aa
 overlap"
 /codon_start=1
 /product="r1fin, MAL1P4.04"
 /protein_id="CAB89212.1"
 /db_xref="GI:7670008"
 misc_feature
 34985..34991
 /note="potential splice donor sequence, aag/gtaata"
 36101..36110
 /note="potential splice acceptor sequence"
 complement(join(39205..40146,40377..40430))
 /gene="r1fin"

Search completed: October 17, 2001, 19:51:13
Job time: 18468 sec

```

/translatatle="MKHITHTLPLKPLNTLHVRHKKPSITSHHIOITRLTCECELE"
/TPANDNPEKXSWOOFHORTJOPREHENDENILKEKROCKXCKOKEIOILIKDIEK
ELTEKPSLOTJDHSHAPICICERSLADKYKRNKLTQNLGKIVAPSSVLAGISE
AALSWKMTETIAAMELAKOQAGAGSHGLAGNVAIEDLRJGTIFVGRILLETI
IDVNWNVSEFIDYDKYSHHTVSCPSLNDVLVTFTMPOFNLVHNSLQGSFYS
SAQVITLIEEVEEVAAGEQATQATVMTPIYTFTEFTAKNKLEVEAATSTSYPIILAS
IAVIVILVIMTVITLIRLRMRKMKIKKKLQYIKLLEE"
50639. .50644
/note="potential splice donor sequence, ttg/gtaaca"
50784. .50793
/note="potential splice acceptor sequence"
complement(53425. .53430)
/note="potential splice donor sequence, aag/gtaaat, at 3'
end of MALP4."
complement(53431. .53502)
/gene="MALP4.05, partial"
complement(53431. .53502)
complement(53431. .53502)
/partial
/gene="MALP4.05, partial"
/note="MALP4.05, partial, len: 24 aa, potential
exon/pseudogene, similarity: PF01045W, 096294 PREMP1
fragment(431 aa) fasta score: 95.8% identity in 24 aa
overlap"
/codon.start=1
/product="potential exon/pseudogene, MALP4.05"
/protein_id="CAB89213.1"
/db_xref="GI:7670009"
/translatatle="MMSISAPLVSQIAPATLSYLLK"
complement(54000. .57115)
/gene="MALP4.06"
/note="r1fin"
complement(join(54000. .55010,55160. .55228))
/gene="r1fin"
/note="MALP4.06, r1fin, len: 359 aa, similarity: most to
097336 PF00300C protein (359 aa), fasta scores: opt: 1549,
E(): 0, 68.5% identity in 359 aa overlap"
/product="r1fin, MALP4.06"
/codon.start=1
/protein_id="CAB89214.1"
/db_xref="GI:7670010"
/translatatle="MKHIAHYKLPLPPTLNTLISYHANKKPSITPHHNRSTSR
LSEYDIESIDSEDEIDSKVEIFERQASRLREYDERQEKQRKKRQKROKNIKITI
HKDKKMEIAKEIGGLRGCGGSGSAGSITLGEVAVATNPKPALDAAITPAALNAG
IVKIAAANAAGIEGALVIAEQQKKGISITGGLTKSFEFTISYKNVSSITAAVNGG

```

| | | | | |
|-----------------------|-----------------|-----------------|-----------|---------------|
| Query Match | 14.68; | Score 52.2; | DB 96; | Length 66441; |
| Best Local Similarity | 49.38; | Pred. No. 5.8; | | |
| Matches 166; | Conservative 0; | Mismatches 168; | Indels 3; | Gaps 14; |

| | | | |
|----|-------|--|-------|
| OY | 14 | tgaagatctttaaactcctctcttttcttgtaaatctaaacattcttttcgcca | 73 |
| Db | 35449 | TTAAAAATTTTTTCCTCTTTTTTTTTTTTTTGTATTATAAATTTTTTTTAT | 35390 |
| OY | 74 | algabataaaaatltag--aaabaataltalcgatatcglltlttagtcggg | 130 |
| Db | 35389 | TTATTTCTAAATAAAATTTTTTTTAATCATTTTTTTTATATAAATTTTTTAAAT | 35330 |
| OY | 131 | tttgagaagaaggttcgacctcgaaagtgagcatgatagatggagagtcggt | 190 |
| Db | 35329 | CTTTTGATAATATTTTTCATTTTTTATTTTATCAAAATTTTATTTTTTAAITTT | 35270 |
| OY | 191 | gagcttcggacatttgtatlgatgcgttgatattagtgcgacaatalaacct | 250 |
| Db | 35269 | TATATTTTTTAAAATATTTTTCTCCTTTTTTTTTTTTTTTTTTAAATAATTT | 35210 |
| OY | 251 | aaagtgcttcctaagaagcccaatlatatcgaattaaacaagtgacaacttact | 310 |
| Db | 35209 | TTTTTTTCTTTTTTCATTTTTTTTCTTTATCAATATAATACATATATATAAATATATA | 35150 |
| OY | 311 | tcglltctgatccgaagcaataaacaattgtcaata | 347 |
| Db | 35149 | TAAACACATATCATATACACAATATATATCCAAACA | 35113 |

| Query Match | 100.0% | Score 358 | DB 21 | Length 6888 |
|-----------------------|---|-------------------|----------|-------------|
| Best Local Similarity | 100.0% | Pred. No. 8.7e-56 | | |
| Matches 358 | Conservative 0 | Mismatches 0 | Indels 0 | Gaps |
| QY 1 | aaaaaaaaaagatgaagatcttctatctctctctttttttgataattcaaca | 60 | | |
| Db 6111 | aaaaaaaaaagatgaagatcttctatctctctctttttttgataattcaaca | 6177 | | |
| QY 61 | tttttttgcctaataatataaatttgataaataatatttgatattcgttt | 120 | | |
| Db 6171 | tttttttgcctaataatataaatttgataaataatatttgatattcgttt | 6230 | | |
| QY 121 | ttatttcgggttgagagaagaaggtcttcgcagatttcgagaattgacgtatgataagattgg | 180 | | |
| Db 6531 | ttatttcgggttgagagaagaaggtcttcgcagatttcgagaattgacgtatgataagattgg | 6290 | | |

```

RESULT 2
AAFS6252/C
ID AAF58252 standard; DNA; 936 BP.
XX
AC AAF58252;
XX
DT 24-APR-2001 (first entry)
XX
DE Oligonucleotide D1835.
XX
KW Electron-transfer group; ETM; mismatch; genotyping;
KW gene expression; ss.
XX
OS Synthetic.
XX
PN WO200107665-A2.
XX
PD 01-FEB-2001.
XX
PE 26-JUL-2000; 2000WO-US20476.
XX
PR 26-JUL-1999; 99US-0145695.
PR 17-MAR-2000; 2000US-0190259.
XX
PA (CLIN-) CLINICAL MICRO SENSORS INC.
XX
PI
XX
PI
XX
DR WPI; 2001-159728/16.
XX
PT Nucleic acids containing electron-transfer group, useful as labels in
PT hybridization assays, e.g. for genotyping, allowing repeat analyses on
PT a single surface.
XX
PS
XX
PS Example 6; Page 127; 159pp; English.
XX
CC The present invention relates to a composition comprising two nucleic
CC acids each containing an electron-transfer group (ETM) having
CC different redox potentials. The invention is used for electronic
CC detection of nucleic acids, especially of substitutions (mismatches)
CC and single-nucleotide polymorphisms, e.g. for genotyping,
CC monitoring gene expression.
XX
SQ Sequence 936 BP; 4 A; 139 C; 10 G; 7 T; 776 other;
XX
Query Match 27.5%; Score 98.6; DB 22; Length 936;
Best Local Similarity 1.4%; Pred. No. 1e-09;
Matches 5; Conservative 254; Mismatches 98; Indels 0; Gaps
OY 1 aaaaaaaaaaagatgaaagtatttcatctctctcttttttttttgataatttaacaa 60
DB 661 wwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwww 602
OY 61 ttttttgcacatataataaaatttgataaataatattgataatcgctt 120
DB 601 wwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwww 542
OY 121 ttagcttgaggttgagaaagagtttcgacttcgaaagtgcagatgatalatagatg 180

```



```
Db 601 WWWW... 542
Oy 121 ttagtcggttgagaaaggttcgacttcgaagtgacatgataatgac 180
Db 541 WWWW... 482
Oy 181 gagcaggttgagcttcggacattgcatgctgattgattagtcgacact 240
Db 481 WWWW... 422
Oy 241 attaaccttaacgtggttcctataagcccaattatacgtataaacaagtgac 300
Db 421 WWWW... 362
Oy 301 aactttacttcgttttgatccgaagcaatacaaatgtcaataacacacaa 357
Db 361 WWWW... 305
```

RESULT 5

```
AAF58259/C
ID AAF58259 standard; DNA; 936 BP.
XX
AC AAF58259;
XX
DT 24-APR-2001 (first entry)
XX
DE Oligonucleotide D2004.
XX
KW Electron-transfer group; ETM; mismatch; genotyping;
KM gene expression; ss.
XX
OS Synthetic.
XX
PN WO200107665-A2.
XX
PD 01-FEB-2001.
XX
PE 26-JUL-2000; 2000WO-US20476.
XX
PR 26-JUL-1999; 99US-0145695.
XX
PR 17-MAR-2000; 2000US-0190259.
XX
PA (CLIN-) CLINICAL MICRO SENSORS INC.
XX
PI Umek RM;
XX
PI WPI; 2001-159728/16.
XX
DR WPI; 2001-159728/16.
XX
PT Nucleic acids containing electron-transfer group, useful as labels in
PT hybridization assays, e.g. for genotyping, allowing repeat analyses on
PT a single surface.
XX
PS Example 6; Page 128; 159pp; English.
XX
CC The present invention relates to a composition comprising two nucleic
CC acids each containing an electron-transfer group (ETM) having
CC different redox potentials. The invention is used for electronic
CC detection of nucleic acids, especially of substitutions (mismatches)
CC and single-nucleotide polymorphisms, e.g. for genotyping,
CC monitoring gene expression.
XX
SO Sequence 936 BP; 6 A; 138 C; 8 G; 8 T; 776 other;
```

```
Query Match 27.5%; Score 98.6; DB 22; Length 936;
Best Local Similarity 1.4%; Pred. No. 1e-09;
Matches 5; Conservative 254; Mismatches 98; Indels 0; Gaps 0;
Oy 1 aaaaaaaaaagatgaagattttattctctctcttttttttgataatttaacaa 60
Db 661 WWWW... 602
```

```
Oy 61 ttttttcccatgatataaaatttgataataataattgatatctgttc 120
Db 601 WWWW... 542
Oy 121 ttagtcggttgagaaaggttcgacttcgaagtgacatgataatgac 180
Db 541 WWWW... 482
Oy 181 gagcaggttgagcttcggacattgcatgctgattgattagtcgacact 240
Db 481 WWWW... 422
Oy 241 attaaccttaacgtggttcctataagcccaattatacgtataaacaagtgac 300
Db 421 WWWW... 362
Oy 301 aactttacttcgttttgatccgaagcaatacaaatgtcaataacacacaa 357
Db 361 WWWW... 305
```

RESULT 6

```
AAF58262/C
ID AAF58262 standard; DNA; 936 BP.
XX
AC AAF58262;
XX
DT 24-APR-2001 (first entry)
XX
DE Oligonucleotide D2007.
XX
KW Electron-transfer group; ETM; mismatch; genotyping;
KM gene expression; ss.
XX
OS Synthetic.
XX
PN WO200107665-A2.
XX
PD 01-FEB-2001.
XX
PE 26-JUL-2000; 2000WO-US20476.
XX
PR 26-JUL-1999; 99US-0145695.
XX
PR 17-MAR-2000; 2000US-0190259.
XX
PA (CLIN-) CLINICAL MICRO SENSORS INC.
XX
PI Umek RM;
XX
PI WPI; 2001-159728/16.
XX
DR WPI; 2001-159728/16.
XX
PT Nucleic acids containing electron-transfer group, useful as labels in
PT hybridization assays, e.g. for genotyping, allowing repeat analyses on
PT a single surface.
XX
PS Example 6; Page 128; 159pp; English.
XX
CC The present invention relates to a composition comprising two nucleic
CC acids each containing an electron-transfer group (ETM) having
CC different redox potentials. The invention is used for electronic
CC detection of nucleic acids, especially of substitutions (mismatches)
CC and single-nucleotide polymorphisms, e.g. for genotyping,
CC monitoring gene expression.
XX
SO Sequence 936 BP; 5 A; 139 C; 10 G; 6 T; 776 other;
```

```
Query Match 27.5%; Score 98.6; DB 22; Length 936;
Best Local Similarity 1.4%; Pred. No. 1e-09;
Matches 5; Conservative 254; Mismatches 98; Indels 0; Gaps 0;
Oy 1 aaaaaaaaaagatgaagattttattctctctcttttttttgataatttaacaa 60
```

```
Db 661 WWWWWW.....:
Oy 61 ttttttgcacatgataataaaattgataataattcgtttt 120
Db 601 WWWWWW.....:
Oy 121 ttatgctgggttggaaaagggttcgacttcgaagtgacagtataatgac 180
Db 541 WWWWWW.....:
Oy 181 gaggcaggttgacgttcgttcgacatttgatgagtgatgacact 240
Db 481 WWWWWW.....:
Oy 241 attaaccttaagtgccttcctataagcccaattatacgtataaacaagtac 300
Db 421 WWWWWW.....:
Oy 301 aactttactcgttttggatccgagcaataacaattgcaataacacacaa 357
Db 361 WWWWWW.....:

RESULT 7
AAF58255/c
ID AAF58255 standard; DNA; 938 BP.
AC AAF58255;
XX
XX
XX 24-APR-2001 (first entry)
DT
DE Oligonucleotide D1876.
XX
XX Electron-transfer group; ETM; mismatch; genotyping;
KW gene expression; ss.
XX
XX Synthetic.
OS
XX WO200107665-A2.
XX
XX 01-FEB-2001.
PD
XX
XX 26-JUL-2000; 2000WO-US20476.
PE
XX
XX 26-JUL-1999; 99US-0145695.
PR
XX 17-MAR-2000; 2000US-0190259.
PA (CLIN-) CLINICAL MICRO SENSORS INC.
XX
XX Umek RM;
PI
XX WPI; 2001-159728/16.
DR
XX
XX Nucleic acids containing electron-transfer group, useful as labels in
PT hybridization assays, e.g. for genotyping, allowing repeat analyses on
XX a single surface
XX
XX Example 6; Page 127; 159pp; English.
XX
XX The present invention relates to a composition comprising two nucleic
CC acids each containing an electron-transfer group (ETM) having
CC different redox potentials. The invention is used for electronic
CC detection of nucleic acids, especially of substitutions (mismatches)
CC and single-nucleotide polymorphisms, e.g. for genotyping,
CC monitoring gene expression.
XX
XX Sequence 938 BP; 4 A; 144 C; 9 G; 5 T; 776 other;
SO
```

Query Match 27.5%; Score 98.6; DB 22; Length 938;
Best Local Similarity 1.4%; Pred. No. 1e-09;
Matches 5; Conservative 254; Mismatches 98; Indels 0; Gaps 0;

```
Oy 1 aaaaaaaaaatgaatattttatctctcttttttttgataatttaaca 60
Db 661 WWWWWW.....:
Oy 61 ttttttgcacatgataataaaattgataataattcgtttt 120
Db 601 WWWWWW.....:
Oy 121 ttatgctgggttggaaaagggttcgacttcgaagtgacagtataatgac 180
Db 541 WWWWWW.....:
Oy 181 gaggcaggttgacgttcgttcgacatttgatgagtgatgacact 240
Db 481 WWWWWW.....:
Oy 241 attaaccttaagtgccttcctataagcccaattatacgtataaacaagtac 300
Db 421 WWWWWW.....:
Oy 301 aactttactcgttttggatccgagcaataacaattgcaataacacacaa 357
Db 361 WWWWWW.....:

RESULT 8
AAF58252
ID AAF58252 standard; DNA; 936 BP.
AC AAF58252;
XX
XX
XX 24-APR-2001 (first entry)
DT
DE Oligonucleotide D1835.
XX
XX Electron-transfer group; ETM; mismatch; genotyping;
KW gene expression; ss.
XX
XX Synthetic.
OS
XX WO200107665-A2.
XX
XX 01-FEB-2001.
PD
XX
XX 26-JUL-2000; 2000WO-US20476.
PE
XX
XX 26-JUL-1999; 99US-0145695.
PR
XX 17-MAR-2000; 2000US-0190259.
PA (CLIN-) CLINICAL MICRO SENSORS INC.
XX
XX Umek RM;
PI
XX WPI; 2001-159728/16.
DR
XX
XX Nucleic acids containing electron-transfer group, useful as labels in
PT hybridization assays, e.g. for genotyping, allowing repeat analyses on
XX a single surface
XX
XX Example 6; Page 127; 159pp; English.
XX
XX The present invention relates to a composition comprising two nucleic
CC acids each containing an electron-transfer group (ETM) having
CC different redox potentials. The invention is used for electronic
CC detection of nucleic acids, especially of substitutions (mismatches)
CC and single-nucleotide polymorphisms, e.g. for genotyping,
CC monitoring gene expression.
XX
XX Sequence 936 BP; 4 A; 139 C; 10 G; 7 T; 776 other;
SO
```

Query Match 27.1%; Score 97; DB 22; Length 936;

Best Local Similarity 1.1%; Pred. No. 1.9e-09;
Matches 4; Conservative 254; Mismatches 99; Indels 0; Gaps 0;

```
Oy 1 aaaaaaaaaagatgaagatatttattctctcttttttttgataatttaacaa 60
Db 324 www. ....
Oy 61 ttttttgcgaatgatatataaaatttgataataatatttgatattcgttt 120
Db 324 www. ....
Oy 121 ttacttcggtttgagaaagggttcgaacttcgaagtgacgaltatagattg 180
Db 444 www. ....
Oy 181 gagcaggttgagctcttgacatttgatgtgattgattatagtgacact 240
Db 504 www. ....
Oy 241 attaaccttaaatggtcttctataagcccaattatatacgattatacaaaagtac 300
Db 564 www. ....
Oy 301 aactttacttcgtttttgatccgaagcaatacaaatgtcaataaccacaa 357
Db 624 www. ....
```

RESULT 9

AAFS8254
ID AAF58254 standard; DNA; 936 BP.

AAFS8254;

24-APR-2001 (first entry)

Oligonucleotide D1875.

Electron-transfer group; ETM; mismatch; genotyping;

gene expression; ss.

Synthetic.

MO200107665-A2.

01-FEB-2001.

26-JUL-2000; 2000MO-US20476.

26-JUL-1999; 99US-0145695.

17-MAR-2000; 2000US-0190259.

(CLIN-) CLINICAL MICRO SENSORS INC.

Umek RM;

WPI: 2001-159728/16.

Nucleic acids containing electron-transfer group, useful as labels in hybridization assays, e.g. for genotyping, allowing repeat analyses on a single surface

Example 6; Page 127; 159pp; English.

The present invention relates to a composition comprising two nucleic acids each containing an electron-transfer group (ETM) having different redox potentials. The invention is used for electronic detection of nucleic acids, especially of substitutions (mismatches) and single-nucleotide polymorphisms, e.g. for genotyping, monitoring gene expression.

Sequence 936 BP; 4 A; 144 C; 7 G; 5 T; 776 other;

Query Match 27.1%; Score 97; DB 22; Length 936;

Best Local Similarity 1.1%; Pred. No. 1.9e-09;
Matches 4; Conservative 254; Mismatches 99; Indels 0; Gaps 0;

```
Oy 1 aaaaaaaaaagatgaagatatttattctctcttttttttgataatttaacaa 60
Db 324 www. ....
Oy 61 ttttttgcgaatgatatataaaatttgataataatatttgatattcgttt 120
Db 324 www. ....
Oy 121 ttacttcggtttgagaaagggttcgaacttcgaagtgacgaltatagattg 180
Db 444 www. ....
Oy 181 gagcaggttgagctcttgacatttgatgtgattgattatagtgacact 240
Db 504 www. ....
Oy 241 attaaccttaaatggtcttctataagcccaattatatacgattatacaaaagtac 300
Db 564 www. ....
Oy 301 aactttacttcgtttttgatccgaagcaatacaaatgtcaataaccacaa 357
Db 624 www. ....
```

RESULT 10

AAFS8257
ID AAF58257 standard; DNA; 936 BP.

AAFS8257;

24-APR-2001 (first entry)

Oligonucleotide D1954.

Electron-transfer group; ETM; mismatch; genotyping;

gene expression; ss.

Synthetic.

MO200107665-A2.

01-FEB-2001.

26-JUL-2000; 2000MO-US20476.

26-JUL-1999; 99US-0145695.

17-MAR-2000; 2000US-0190259.

(CLIN-) CLINICAL MICRO SENSORS INC.

Umek RM;

WPI: 2001-159728/16.

Nucleic acids containing electron-transfer group, useful as labels in hybridization assays, e.g. for genotyping, allowing repeat analyses on a single surface

Example 6; Page 127; 159pp; English.

The present invention relates to a composition comprising two nucleic acids each containing an electron-transfer group (ETM) having different redox potentials. The invention is used for electronic detection of nucleic acids, especially of substitutions (mismatches) and single-nucleotide polymorphisms, e.g. for genotyping, monitoring gene expression.

Sequence 936 BP; 5 A; 142 C; 7 G; 6 T; 776 other;

Query Match 27.1%; Score 97; DB 22; Length 936;

Best Local Similarity 1.1%; Pred. No. 1.9e-09; Mismatches 99; Indels 0; Gaps 0;

Matches 4; Conservative 254; Mismatches 99; Indels 0; Gaps 0;

OY 1 aaaaaaaaaaagatgaagatttattctctctcttttttttgataatttaataca 60
 Db 324 www..... 383
 OY 61 ttttttgcacatgatatataaaatttgataaataattgataattcgtttt 120
 Db 384 www..... 443
 OY 121 ttagtcggttgagaaaaggcttcgacttcgaaattgagcagatgatatagattgg 180
 Db 444 www..... 503
 OY 181 gagctaggtgagcttcgacattgtgattgagtggttgattagtgctgcacact 240
 Db 504 www..... 563
 OY 241 ataaaccttaaatgggcttcctataaggcccaattatatacgaataaagtgac 300
 Db 564 www..... 623
 OY 301 aactttactcgttttgcgcgaagcaataaattgtccaataccaacacaa 357
 Db 624 www..... 680

RESULT 11

ID AAF58259 standard; DNA; 936 BP.

AC AAF58259;

DT 24-APR-2001 (first entry)

DE Oligonucleotide D2004.

KW Electron-transfer group; ETW; mismatch; genotyping;
 gene expression; ss.

OS Synthetic.

PN WO200107665-A2.

PD 01-FEB-2001.

PF 26-JUL-2000; 2000WO-US20476.

PR 26-JUL-1999; 99US-0145695.

PR 17-MAR-2000; 2000US-0190259.

PA (CLIN-) CLINICAL MICRO SENSORS INC.

PI umek RM;

DR WPI; 2001-159728/16.

PT Nucleic acids containing electron-transfer group, useful as labels in
 hybridization assays, e.g. for genotyping, allowing repeat analyses on
 a single surface

PS Example 6; Page 128; 159pp; English.

CC The present invention relates to a composition comprising two nucleic
 acids each containing an electron-transfer group (ETW) having
 different redox potentials. The invention is used for electronic
 detection of nucleic acids, especially of substitutions (mismatches)
 and single-nucleotide polymorphisms, e.g. for genotyping,

CC monitoring gene expression.

Sequence 936 BP; 6 A; 138 C; 8 G; 8 T; 776 other;

Query Match 27.1%; Score 97; DB 22; Length 936;

Best Local Similarity 1.1%; Pred. No. 1.9e-09; Mismatches 99; Indels 0; Gaps 0;

Matches 4; Conservative 254; Mismatches 99; Indels 0; Gaps 0;

OY 1 aaaaaaaaaaagatgaagatttattctctctcttttttttgataatttaataca 60
 Db 324 www..... 383
 OY 61 ttttttgcacatgatatataaaatttgataaataattgataattcgtttt 120
 Db 384 www..... 443
 OY 121 ttagtcggttgagaaaaggcttcgacttcgaaattgagcagatgatatagattgg 180
 Db 444 www..... 503
 OY 181 gagctaggttgagcttcgacattgtgattgagtggttgattagtgctgcacact 240
 Db 504 www..... 563
 OY 241 ataaaccttaaatgggcttcctataaggcccaattatatacgaataaagtgac 300
 Db 564 www..... 623
 OY 301 aactttactcgttttgcgcgaagcaataaattgtccaataccaacacaa 357
 Db 624 www..... 680

RESULT 12

ID AAF58262 standard; DNA; 936 BP.

AC AAF58262;

DT 24-APR-2001 (first entry)

DE Oligonucleotide D2007.

KW Electron-transfer group; ETW; mismatch; genotyping;
 gene expression; ss.

OS Synthetic.

PN WO200107665-A2.

PD 01-FEB-2001.

PF 26-JUL-2000; 2000WO-US20476.

PR 26-JUL-1999; 99US-0145695.

PR 17-MAR-2000; 2000US-0190259.

PA (CLIN-) CLINICAL MICRO SENSORS INC.

PI umek RM;

DR WPI; 2001-159728/16.

PT Nucleic acids containing electron-transfer group, useful as labels in
 hybridization assays, e.g. for genotyping, allowing repeat analyses on
 a single surface

PS Example 6; Page 128; 159pp; English.

CC The present invention relates to a composition comprising two nucleic
 acids each containing an electron-transfer group (ETW) having
 different redox potentials. The invention is used for electronic
 detection of nucleic acids, especially of substitutions (mismatches)
 and single-nucleotide polymorphisms, e.g. for genotyping,

CC detection of nucleic acids, especially of substitutions (mismatches)
CC and single-nucleotide polymorphisms, e.g. for genotyping,
CC monitoring gene expression.
XX
SQ Sequence 936 BP; 5 A; 139 C; 10 G; 6 T; 776 other;

Query Match 27.1%; Score 97; DB 22; Length 936;
Best Local Similarity 1.1%; Pred. No. 1.9e-09;
Matches 4; Conservative 254; Mismatches 99; Indels 0; Gaps 0;

QY 1 aaaaaaaaaaagatgaagatttattctctctcttttttttgataatttaaca 60
DB 324 www.
QY 61 ttttttcccaatgatatataaaatttgataataatattatgtatctgttt 120
DB 384 www.
QY 121 ttatgtcgggtttgagaaaagggttcgacttcgaaagtgacgtatataagttg 180
DB 444 www.
QY 181 gaggcaggttgaggtcttgacattgtatcgatgtgttgattatagtgcgacact 240
DB 504 www.
QY 241 attaaacctaaatgggcttctataagcccaattatatacgattataacaagtgc 300
DB 564 www.
QY 301 aactttactcgttttgcgcgaagcaataacaaatgtcaatccaacaacaa 357
DB 624 www.

RESULT 13

AAF58255 standard; DNA; 938 BP.

AAF58255;

24-APR-2001 (first entry)

Oligonucleotide D1876.

Electron-transfer group; ETM; mismatch; genotyping;

gene expression; ss.

Synthetic.

WO200107665-A2.

01-FEB-2001.

26-JUL-2000; 2000WO-US20476.

26-JUL-1999; 99US-0145695.

17-MAR-2000; 2000US-0190259.

(CLIN-) CLINICAL MICRO SENSORS INC.

Umek RM;

WPI; 2001-159728/16.

Nucleic acids containing electron-transfer group, useful as labels in
hybridization assays, e.g. for genotyping, allowing repeat analyses on
a single surface

Example 6; Page 127; 159pp; English.

The present invention relates to a composition comprising two nucleic

CC acids each containing an electron-transfer group (ETM) having
CC different redox potentials. The invention is used for electronic
CC detection of nucleic acids. Especially of substitutions (mismatches)
CC and single-nucleotide polymorphisms, e.g. for genotyping,
CC monitoring gene expression.
XX
SQ Sequence 938 BP; 4 A; 144 C; 9 G; 5 T; 776 other;

Query Match 27.1%; Score 97; DB 22; Length 938;
Best Local Similarity 1.1%; Pred. No. 1.9e-09;
Matches 4; Conservative 254; Mismatches 99; Indels 0; Gaps 0;

QY 1 aaaaaaaaaaagatgaagatttattctctctcttttttttgataatttaaca 60
DB 324 www.
QY 61 ttttttcccaatgatatataaaatttgataataatattatgtatctgttt 120
DB 384 www.
QY 121 ttatgtcgggtttgagaaaagggttcgacttcgaaagtgacgtatataagttg 180
DB 444 www.
QY 181 gaggcaggttgaggtcttgacattgtatcgatgtgttgattatagtgcgacact 240
DB 504 www.
QY 241 attaaacctaaatgggcttctataagcccaattatatacgattataacaagtgc 300
DB 564 www.
QY 301 aactttactcgttttgcgcgaagcaataacaaatgtcaatccaacaacaa 357
DB 624 www.

RESULT 14

AAF58238/C standard; DNA; 244 BP.

AAF58238;

24-APR-2001 (first entry)

Oligonucleotide D1250:D1102.

Electron-transfer group; ETM; mismatch; genotyping;

gene expression; ss.

Synthetic.

WO200107665-A2.

01-FEB-2001.

26-JUL-2000; 2000WO-US20476.

26-JUL-1999; 99US-0145695.

17-MAR-2000; 2000US-0190259.

(CLIN-) CLINICAL MICRO SENSORS INC.

Umek RM;

WPI; 2001-159728/16.

Nucleic acids containing electron-transfer group, useful as labels in
hybridization assays, e.g. for genotyping, allowing repeat analyses on
a single surface

Example 4; Page 120; 159pp; English.

```
XX The present invention relates to a composition comprising two nucleic
CC acids each containing an electron-transfer group (ETM) having
CC different redox potentials. The invention is used for electronic
CC detection of nucleic acids, especially of substitutions (mismatches)
CC and single-nucleotide polymorphisms, e.g. for genotyping,
CC monitoring gene expression.
XX
SQ Sequence 244 BP; 19 A; 9 C; 12 G; 10 T; 194 other;

Query Match      16.5%; Score 59.2; DB 22; Length 244;
Best Local Similarity 0.6%; Pred.No. 0.01;
Matches 1; Conservative 137; Mismatches 40; Indels 0; Gaps 0;

QY 1 aaaaaaaaaagatgaagatatttattctctctcttttttttgataatttaataca 60
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 196 wwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwww 137
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 61 ttttttgcacatgatatataaaatttgataataatttgatatattcgttt 120
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 136 wwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwww 77
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 121 ttatgtcggttgagaaaagggttcgacttcgaagtgagacgatagataagatt 178
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 76 wwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwww 19

RESULT 15
AAF58238
ID AAF58238 standard; DNA; 244 BP.
XX
AC AAF58238;
XX
DT 24-APR-2001 (first entry)
XX
DE Oligonucleotide D1250:D1102.
XX
KM Electron-transfer group: ETM; mismatch; genotyping;
XX gene expression; ss.
XX
OS Synthetic.
XX
PN WO200107665-A2.
XX
PD 01-FEB-2001.
XX
PF 26-JUL-2000; 2000WO-US20476.
XX
PR 26-JUL-1999; 98US-0145695.
XX 17-MAR-2000; 2000US-0190259.
XX
PA (CLIN-) CLINICAL MICRO SENSORS INC.
XX
PI Umek RM;
XX
DR WPI; 2001-159728/16.
XX
PT Nucleic acids containing electron-transfer group, useful as labels in
PT hybridization assays, e.g. for genotyping, allowing repeat analyses on
PT a single surface
XX
PS Example 4; Page 120; 159pp; English.
XX
CC The present invention relates to a composition comprising two nucleic
CC acids each containing an electron-transfer group (ETM) having
CC different redox potentials. The invention is used for electronic
CC detection of nucleic acids, especially of substitutions (mismatches)
CC and single-nucleotide polymorphisms, e.g. for genotyping,
CC monitoring gene expression.
XX
SQ Sequence 244 BP; 19 A; 9 C; 12 G; 10 T; 194 other;
```

```
Query Match      16.2%; Score 58; DB 22; Length 244;
Best Local Similarity 4.0%; Pred.No. 0.017;
Matches 7; Conservative 127; Mismatches 42; Indels 0; Gaps 0;

QY 1 aaaaaaaaaagatgaagatatttattctctctcttttttttgataatttaataca 60
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 37 wwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwww 96
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 61 ttttttgcacatgatatataaaatttgataataatttgatatattcgttt 120
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 97 wwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwww 156
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 121 ttatgtcggttgagaaaagggttcgacttcgaagtgagacgatagataaga 176
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 157 wwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwww 212

Search completed: October 17, 2001, 19:19:16
Job time: 15956 sec
```


GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 17, 2001, 19:14:07 ; Search time 156 Seconds
(without alignments)
434.445 Million cell updates/sec

Title: US-09-502-426-1_COPY_6111_6468
Perfect score: 358
Sequence: 1 aaaaaaaaaatgaagt.....tgcataatcacaacacaaag 358

Scoring table: IDENTITY_NIC
Gapop 10.0 , Gapext 1.0

Searched: 324599 seqs, 9465562 residues

Total number of hits satisfying chosen parameters: 649198

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents_NA:*
1: /cgn2_6/prodata/2/lna/5A_COMB.seq:*
2: /cgn2_6/prodata/2/lna/5B_COMB.seq:*
3: /cgn2_6/prodata/2/lna/6A_COMB.seq:*
4: /cgn2_6/prodata/2/lna/6B_COMB.seq:*
5: /cgn2_6/prodata/2/lna/PCTUS_COMB.seq:*
6: /cgn2_6/prodata/2/lna/Backfilest1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description |
|------------|-------|-------------|--------|----|-------------------|
| 1 | 42.8 | 12.0 | 1493 | 1 | US-08-340-820-24 |
| 2 | 42.8 | 12.0 | 1493 | 1 | US-08-593-535-24 |
| 3 | 41.8 | 11.7 | 998 | 4 | US-09-122-4008-5 |
| 4 | 41 | 11.5 | 2230 | 4 | US-08-378-313-24 |
| 5 | 40.8 | 11.4 | 6243 | 2 | US-09-056-075-1 |
| 6 | 40.8 | 11.4 | 19557 | 5 | PCT-US92-06300-1 |
| 7 | 40.6 | 11.3 | 1733 | 3 | US-09-073-569-1 |
| 8 | 40.6 | 11.3 | 3095 | 6 | 5231168-1 |
| 9 | 40.4 | 11.3 | 10607 | 1 | US-08-078-090-3 |
| 10 | 40 | 11.2 | 2755 | 3 | US-08-749-522-2 |
| 11 | 39.4 | 11.0 | 289 | 1 | US-08-341-568-3 |
| 12 | 39.4 | 11.0 | 289 | 2 | US-08-911-020-3 |
| 13 | 39.4 | 11.0 | 19124 | 2 | US-08-487-826B-13 |
| 14 | 39 | 10.9 | 9048 | 3 | US-08-973-273-4 |
| 15 | 38.8 | 10.8 | 835 | 4 | US-08-998-416-547 |
| 16 | 38.8 | 10.8 | 863 | 4 | US-08-998-416-498 |
| 17 | 38.8 | 10.8 | 3138 | 1 | US-07-867-106-4 |
| 18 | 38.2 | 10.7 | 2836 | 3 | US-08-747-221B-24 |
| 19 | 38.2 | 10.7 | 2836 | 3 | US-08-747-221B-26 |
| 20 | 38.2 | 10.7 | 25604 | 1 | US-07-867-106-2 |
| 21 | 38.2 | 10.7 | 5852 | 3 | US-08-747-221B-24 |
| 22 | 38 | 10.6 | 701 | 4 | US-08-998-416-701 |
| 23 | 38 | 10.6 | 2555 | 2 | US-08-693-457-3 |
| 24 | 38 | 10.6 | 2555 | 4 | US-09-265-731-3 |
| 25 | 38 | 10.6 | 8700 | 2 | US-08-393-625-16 |
| 26 | 38 | 10.6 | 8700 | 2 | US-08-466-961A-16 |
| 27 | 38 | 10.6 | 8700 | 2 | US-08-645-193B-18 |

| | | | | | | |
|----|------|------|------|---|--------------------|--------------------|
| 28 | 37.8 | 10.6 | 1582 | 3 | US-08-545-196B-10 | Sequence 10, Appl |
| 29 | 37.8 | 10.6 | 1582 | 3 | US-08-545-196B-12 | Sequence 12, Appl |
| 30 | 37.6 | 10.5 | 2422 | 1 | US-07-867-106-5 | Sequence 5, Appl |
| 31 | 37.6 | 10.5 | 3138 | 1 | US-07-867-106-4 | Sequence 4, Appl |
| 32 | 37.6 | 10.5 | 5852 | 1 | US-07-867-106-2 | Sequence 2, Appl |
| 33 | 37.6 | 10.5 | 6078 | 4 | US-09-173-914-1 | Sequence 1, Appl |
| 34 | 37.6 | 10.4 | 2445 | 6 | 5215909-9 | Patent No. 5215909 |
| 35 | 37.2 | 10.4 | 144 | 1 | US-08-702-344-26 | Sequence 26, Appl |
| 36 | 37.2 | 10.4 | 240 | 1 | US-08-628-417-6 | Sequence 186, App |
| 37 | 37.2 | 10.4 | 615 | 4 | US-08-998-416-186 | Sequence 186, App |
| 38 | 37.2 | 10.4 | 837 | 4 | US-08-998-416-1137 | Sequence 1137, App |
| 39 | 37.2 | 10.4 | 860 | 1 | US-07-847-010-18 | Sequence 288, App |
| 40 | 37.2 | 10.4 | 1615 | 1 | US-08-413-118-108 | Sequence 108, App |
| 41 | 37.2 | 10.4 | 1615 | 2 | US-08-413-118-108 | Sequence 98, App |
| 42 | 37.2 | 10.4 | 1615 | 2 | US-08-473-446-108 | Sequence 108, App |
| 43 | 37.2 | 10.4 | 2058 | 2 | US-08-749-391-1 | Sequence 1, Appl |
| 44 | 37.2 | 10.4 | 2058 | 2 | US-08-749-391-1 | Sequence 1, Appl |
| 45 | 37.2 | 10.4 | 2058 | 3 | US-09-390-200-1 | Sequence 1, Appl |

ALIGNMENTS

RESULT 1
US-08-340-820-24/c
Sequence 24, Application US/08340820
Patent No. 5512460
GENERAL INFORMATION:
APPLICANT: NARUO, Ken-ichi
APPLICANT: SEKO, Chisao
APPLICANT: KUROKAWA, Tsutomu
APPLICANT: KONDO, Tatsuya
TITLE OF INVENTION: GLIA ACTIVATING FACTOR AND ITS
NUMBER OF SEQUENCES: 27
CORRESPONDENCE ADDRESS:
ADDRESSEE: DAVID G. CONLIN, DIKE, BRONSTEIN, ROBERTS &
ADDRESS: 130 Water Street
CITY: Boston
STATE: Massachusetts
COUNTRY: US
ZIP: 02109
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/340,820
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/07/835,713
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: CONLIN, David G.
REGISTRATION NUMBER: 27026
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617)523-3400
TELEFAX: (617)523-6440
TELEX: 200291 STRE UR
INFORMATION FOR SEQ ID NO: 24:
SEQUENCE CHARACTERISTICS:
LENGTH: 1493 base pairs
TYPE: nucleic acid
STRANDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: cDNA to mRNA
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Homo sapiens


```

US-08-378-313-24/C
; Sequence 24, Application US/08378313
; Patent No. 6207881
; GENERAL INFORMATION:
; APPLICANT: THEOLOGIS, ATHANASIOS
; APPLICANT: SATO, TAKAHITO
; TITLE OF INVENTION: CONTROL OF FRUIT RIPENING THROUGH
; TITLE OF INVENTION: GENETIC CONTROL OF ACC SYNTHASE SYNTHESIS
; NUMBER OF SEQUENCES: 34
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MORRISON & FOEISTER
; STREET: 755 Page Mill Road
; CITY: Palo Alto
; STATE: California
; COUNTRY: USA
; ZIP: 94304-1018
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/378,313
; FILING DATE:
; CLASSIFICATION: 800
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/862,493
; FILING DATE: 02-APR-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: MURASHIGE, KATE H.
; REGISTRATION NUMBER: 29,959
; REFERENCE/DOCKET NUMBER: 29190-20002.20
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 856-5600
; TELEFAX: (415) 494-0792
; TELEX: 706141
; INFORMATION FOR SEQ. ID NO: 24:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2230 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 91..1545
; US-08-378-313-24

```

```

Query Match      11.5%; Score 41; DB 4; Length 2230;
Best Local Similarity 44.9%; Pred. No. 0.85;
Matches 155; Conservative 0; Mismatches 190; Indels 0; Gaps 0;

QY 4 aaaaaaaaaaagaatatttattctctctcttttttttgataaatttaacattt 63
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 2222 AACAAATATATATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTT 2163
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 64 ttttggcccaagataataaaatttgataaaataattatggatattcgtttlla 123
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 2162 TTTTATTTTATTTGCTAAAGCAAAATTAAGTCTAAGTACATTTTATTTTATTTT 2103
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 124 gtctgggtttggaagaaggtttgactttggaagtgagcagatataatagtgagg 183
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 2102 TTAACATACTACCTTTGCTTTGCTTATTAATTAATTAATTAATTAATTAATTA 2043
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 184 ctaggttgagcttctgacatttgatctgagtgctgattatagtgctgacacatt 243
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 2042 CATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTT 1983
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 244 aaaccttaaaagggtcttctataaaggcccaattatatacgaattatacaaaagtgag 303
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 1982 ACAAAATTAAGATGCTAAGTACATTTTATTTTATTTTATTTTATTTTATTTTATTT 1923
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 304 ttctactcgcttttgatccgaagaacaataaattgtcaataac 348

```

```

DB 1922 TTTGCAATCTATTAATTAGCTAAGACAAATTAATGATGCTAAGTAC 1878
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
RESULT 5
US-09-056-075-1/C
; Sequence 1, Application US/09056075
; Patent No. 595368
; GENERAL INFORMATION:
; APPLICANT: Johnson, Eric A.
; APPLICANT: Bradshaw, Marite
; APPLICANT: Rood, Julian
; TITLE OF INVENTION: Expression System for Clostridium
; TITLE OF INVENTION: Species
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Quarles & Brady
; STREET: 1 South Pinckney Street
; CITY: Madison
; STATE: WI
; COUNTRY: US
; ZIP: 53701-2113
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/056,075
; FILING DATE:
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Seay, Nicholas J.
; REGISTRATION NUMBER: 27386
; REFERENCE/DOCKET NUMBER: 960296.95238
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 608-251-5000
; TELEFAX: 608-251-9166
; INFORMATION FOR SEQ. ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 6243 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: misc-feature
; LOCATION: 3770..4013
; OTHER INFORMATION: /note="RP4 origin of DNA transfer (orit) from
; OTHER INFORMATION: plasmid Rp4"
; US-09-056-075-1

```

```

Query Match      11.4%; Score 40.8; DB 2; Length 6243;
Best Local Similarity 47.3%; Pred. No. 1;
Matches 155; Conservative 0; Mismatches 172; Indels 1; Gaps 1;

QY 4 aaaaaaaaaaagaatatttattctctctcttttttttgataaatttaacattt 63
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 1416 AAAATTAATTAATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTT 1257
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 64 ttttggcccaagataataaaatttgataaataattatggatattcgtttlla 122
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 1356 TTTTATTAATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTT 1297
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 123 agtcgggtttggaagaaggtttgactttggaagtgagcagatataatagtgagg 182
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 1296 TTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTT 1237
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 183 gctaggttgagcttctgacatttgatctgagtgctgattatagtgctgacacatt 242
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 1236 TTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTT 1177
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

```

QY 243 taaccttaaaagggccttctataaagcccaattatattacgatatatacaaaagtgacaa 302
DB 1176 ACCGCTGATATAGGAGGACCTTGACCTGCTTTTGGGAGGCTTGAAGATTA 1117
QY 303 ctttaactcgttttgcacgaagcaa 330
DB 1116 TTTTACTTATTAGTTAGTCCGAGGAA 1089

RESULT 6

PCT-US92-06300-1/C
Sequence 1, Application PC/TUS9206300
GENERAL INFORMATION:
APPLICANT: Hurwitz, David R
APPLICANT: Nathan, Margaret
APPLICANT: Shaul, Moshe
TITLE OF INVENTION: Transgenic Protein Production
NUMBER OF SEQUENCES: 1
CORRESPONDENCE ADDRESS:
ADDRESSEE: Rhone-Poulenc Rorer, Inc.
STREET: 500 Virginia Ave., Bldg. 3A
CITY: Ft. Washington
STATE: Pennsylvania
COUNTRY: USA
ZIP: 19034
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US92/06300
FILING DATE: 19920730
CLASSIFICATION: 800
ATTORNEY/AGENT INFORMATION:
NAME: Goodman, Rosanne
REGISTRATION NUMBER: 52,534
REFERENCE/DOCKET NUMBER: A0856-US
TELECOMMUNICATION INFORMATION:
TELEPHONE: (215) 962-4130
TELEFAX: (215) 962-4107
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 19557 base pairs
TYPE: NUCLEIC ACID
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: NO
PUBLICATION INFORMATION:
AUTHORS: Minohetti, P P
AUTHORS: Ruffner, D E
AUTHORS: Kuang, W.-J.
AUTHORS: Dennison, O E
AUTHORS: Hawkins, J W
AUTHORS: Beatlie, W G
AUTHORS: Dugalczyk, A
TITLE: MOLECULAR STRUCTURE OF THE HUMAN ALBUMIN GENE
TITLE: IS REVEALED BY NUCLEOTIDE SEQUENCE WITHIN 011-22
TITLE: OF CHROMOSOME 4
JOURNAL: J. Biol. Chem.
VOLUME: 261
PAGES: 6747-6757
DATE: 1986
RELEVANT RESIDUES IN SEQ ID NO: 1: FROM 1 TO 19002
PCT-US92-06300-1

Query Match 11.4%; Score 40.8; DB 5; Length 19557;
Best Local Similarity 52.3%; Pred. No. 1.1;
Matches 90; Conservative 0; Mismatches 82; Indels 0; Gaps 0;

QY 2 aaaaaaaagaaagatatttctctctcttcttttttgataaatttaacat 61
DB 423 ATATCATATATATGATATATATATTCATTCCTTTTCTTTTCTTTTCTTTT 364
QY 62 tttttcccaatgatatataaaatttgataaataattatgatalcgtttc 121
DB 363 TTTTCTTTTATATATCTTCGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTG 304
QY 122 tagtcgggttgaaagaggttcgacttcgaaagtgacagatata 173
DB 303 TGTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTG 252

RESULT 7

US-09-073-569-1/C
Sequence 1, Application US/09073569
Patent No. 6084088
GENERAL INFORMATION:
APPLICANT: Sheppard, Paul O.
APPLICANT: Grossmann, Angelika
TITLE OF INVENTION: NOVEL TUMOR ANTIGENS
NUMBER OF SEQUENCES: 15
CORRESPONDENCE ADDRESS:
ADDRESSEE: Zymogenetics, Inc.
STREET: 1201 Eastlake Avenue East
CITY: Seattle
STATE: WA
COUNTRY: USA
ZIP: 98102
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/073,569
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Sawislak, Deborah A
REGISTRATION NUMBER: 37,438
REFERENCE/DOCKET NUMBER: 97-14
TELECOMMUNICATION INFORMATION:
TELEPHONE: 206-442-6672
TELEFAX: 206-442-6678
TELEX:
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 1733 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
FEATURE:
NAME/KEY: Coding Sequence
LOCATION: 34..1344
OTHER INFORMATION:
US-09-073-569-1

Query Match 11.3%; Score 40.6; DB 3; Length 1733;
Best Local Similarity 56.3%; Pred. No. 1;
Matches 76; Conservative 0; Mismatches 59; Indels 0; Gaps 0;

QY 20 tatcttattctctcttttttttgataaatttaaacatctttttgccaatgala 79
DB 1715 TTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTT 1656
QY 80 tataaaaattgataaataattatgatalcgttttctgctcggttcgagaaa 139
DB 1655 TTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTT 1596


```

; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
US-08-749-522-2

```

| Query Match | 11.2% | Score 40 | DB 3 | Length 2755 |
|-----------------------|---|---------------|----------------|-------------|
| Best Local Similarity | 48.3% | Pred. No. 1.4 | | |
| Matches 171 | Conservative | 0 | Mismatches 180 | Indels 3 |
| | | | Gaps 2 | |
| QY 6 | aaaaaagatgaagatctttatctctctctcttttttgataatttaacatctt | 65 | | |
| Db 2046 | ATATACTAGTAATGTATTATTTGACCTTTTAAATCGAATTATGTGTATTGCG | 2105 | | |
| QY 66 | tttgcaccaatgatatataaaaaatttggataaaataattatctgatatcgtttttagt | 125 | | |
| Db 2106 | TTTATACCAATTCATCAATTAATAATTTTATTTATTAATTAATTAATGCGTACTCACGAGTGTGG | 2165 | | |
| QY 126 | tcgggttttgagaagaaggtcttcga-cttcggaagtgagcatgatatagatttggagc | 184 | | |
| Db 2166 | TGCACTAAAATTTATTTAAACAATATATTTTGGAAAAATTGATTAATAATCTAATATAGGTT | 2225 | | |
| QY 185 | taagtttgagctcttggacatttgtatctgtaatgagttgttattatagtttgcgacattat | 244 | | |
| Db 2226 | TTGGTTGAATAGTAAGATATATATTATTTCAAAATTAATAATATGTGAGTTTCAAAATCTATTC | 2285 | | |
| QY 245 | aaacttaaatgggcttcttataaaggccaattatattcgaattataacaagaatggacaact | 304 | | |
| Db 2286 | ATGCTATATATTTGG--TACTATATATCTATATAAAATGGATACCTTATAAAGTATCTAAT | 2343 | | |
| QY 305 | tttactctgcttttggatccgaagcaataaacaattgtccaattaccacaacaag | 358 | | |
| Db 2344 | TTAGTTTATGGTTGATGTGATGATATATCCCAATTTATTAATAATAATATATAG | 2397 | | |

RESULT 11
 US-08-341-568-3/c
 Sequence 3, Application US/08341568
 Patent No. 5661021
 GENERAL INFORMATION:
 APPLICANT: Buchert, Johanna
 APPLICANT: Silka-aho, Matti
 APPLICANT: Viikari, Liisa
 APPLICANT: Penttila, Merja
 APPLICANT: Saloheimo, Anu
 APPLICANT: Marjatta, Rannu
 TITLE OF INVENTION: Mannanase enzymes, genes coding for them,
 TITLE OF INVENTION: methods for isolating the genes, and methods for bleaching lignin
 TITLE OF INVENTION: pulps
 NUMBER OF SEQUENCES: 7
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Birch, Stewart, Kolasch and Birch
 STREET: PO Box 747
 CITY: Falls Church
 STATE: VA
 COUNTRY: US
 ZIP: 22040-0747
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: PatentIn Release #1.0, Version #1.30
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/341,568
 FILING DATE: 22-NOV-1994
 CLASSIFICATION: 435
 ATTORNEY/AGENT INFORMATION:
 NAME: Murphy Jr, Gerald M
 REGISTRATION NUMBER: 28,977
 REFERENCE/DOCKET NUMBER: 365-262P
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (703) 205-8000
 TELEFAX: (703) 205-8050
 TELEX: 248345

```

? INFORMATION FOR SEQ ID NO: 3:
?
? SEQUENCE CHARACTERISTICS:
?     LENGTH: 289 base pairs
?     TYPE: nucleic acid
?     STRANDEDNESS: single
?     TOPOLOGY: linear
?
? MOLECULE TYPE: CDNA
?
? ORIGINAL SOURCE:
?
? ORGANISM: Trichoderma reesei
?     STRAIN: QM9414
?
US-08-341-568-3
```

| | Query Match | 11.0%; | Score 39.4; | DB 1; | Length 289; |
|----|-----------------------|---|---|---|-------------------------------------|
| | Best Local Similarity | 67.9%; | Pred. No. 1.6; | | |
| | Matches 55; | Conservative 0; | Mismatches 26; | Indels 0; | Gaps 0; |
| QY | 20 | tattttaa | ctctctctcttcttcttgcga | aaatttcaacatcttcttgc | ccaatgata 79 |
| | | | | | |
| | | | | | |
| Db | 277 | TTT | TTT | TTT | TTTTTTTTTTTTTTTTTTTTTTTATGTGATA 218 |
| QY | 80 | tataaaaa | attcgataa | aaataa 100 | |
| | | | | | |
| Db | 217 | TATATATATATATTCATCA | CAAAAA 197 | | |

RESULT 12
 US-08-911-020-3/c
 : Sequence 3, Application US/08911020
 : Patent No. 5854047
 :
 : GENERAL INFORMATION:
 : APPLICANT: Buchert, Johanna
 : APPLICANT: Slika-aho, Matti
 : APPLICANT: Viikari, Liisa
 : APPLICANT: Penttila, Merja
 : APPLICANT: Saloheimo, Anu
 : APPLICANT: Marjatta, Renua
 : TITLE OF INVENTION: Mannanase enzymes, genes coding for them,
 : TITLE OF INVENTION: methods for isolating the genes, and methods for bleaching
 : NUMBER OF INVENTION: lignocellulosic pulps
 : NUMBER OF SEQUENCES: 7
 : CORRESPONDENCE ADDRESS:
 : ADDRESSEE: Birch, Stewart, Kolasch and Birch
 : STREET: PO Box 747
 : CITY: Falls Church
 : STATE: VA
 : COUNTRY: US
 : ZIP: 22040-0747
 :
 : COMPUTER READABLE FORM:
 : MEDIUM TYPE: Floppy disk
 : COMPUTER: IBM PC compatible
 : OPERATING SYSTEM: PC-DOS/MS-DOS
 : SOFTWARE: Patentin Release #1.0, Version #1.30
 :
 : CURRENT APPLICATION DATA:
 : APPLICATION NUMBER: US/08/911,020
 : FILING DATE: 13-AUG-1997
 : CLASSIFICATION: 435
 :
 : PRIOR APPLICATION DATA:
 : APPLICATION NUMBER: US 08/341,568
 : FILING DATE: 22-NOV-1994
 : ATTORNEY/AGENT INFORMATION:
 : NAME: Murphy Jr, Gerald M
 : REGISTRATION NUMBER: 28,977
 : REFERENCE/DOCKET NUMBER: 365-262P
 : TELECOMMUNICATION INFORMATION:
 : TELEPHONE: (703) 205-8000
 : TELEFAX: (703) 205-8050
 : TELEX: 248345
 :
 : INFORMATION FOR SEQ ID NO: 3:
 : SEQUENCE CHARACTERISTICS:
 : LENGTH: 289 base pairs
 : TYPE: nucleic acid
 : STRANDEDNESS: single

| | | | | | | | |
|---------------------------|--------|-------|----------|----|----|--------|--------|
| Query Match | 11.08; | Score | 39.4; | DB | 2; | Length | 19124; |
| Best Local Similarity | 46.78; | Pred. | No. 2.1; | | | | |
| Matches 161; Conservative | | | | | | Indels | 3; |
| | | | | | | Gaps | 1. |

MOLECULE LIFE: DNA (genomic)
HYPOTHETICAL: NO

```

; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Brassica
US-08-973-273-4

```

| | | | | |
|-----------------------|-------|--------------|------|---------------|
| Query Match | 10.9% | Score 39 | DB 3 | Length 9048 |
| Best Local Similarity | 55.6% | Pred. No. | 2.4 | |
| Matches | 75 | Conservative | 0 | Mismatches 60 |
| | | | | Indels 0 |
| | | | | Gaps 0 |

0Y 6 aaaaagtgtaagtatttttatcttccttctttttttttgataaatttaacaaatcctt 65
 | | | | | | | | | | | | |
Db 347 AAAAAATTTAATAATAGGTTAATCTATATATATTGTGGCACTCGAAACAATATTT 288
0Y 66 ttggcccatgatatafaaaatttgataaataatatatttgatatactgcgttttagt 129

| | | | |
|----|-----|---|-----|
| Db | 287 | TTATTATTAAGTTAAAAAACATTATTATATAAGATATTTCCTAACTATTTCCTAAGTAT | 228 |
| QY | 126 | tcgggtttgagaaaa | 140 |
| | | | |
| Db | 227 | GAGTGTTTTTAAAAA | 213 |

RESULT 15
US-08-998-416-547
; Sequence 547, Application US/08998416
; Patent No. 6239264

; GENERAL INFORMATION:

APPLICANT: Philippson, Peter
APPLICANT: Pohlmann, Rainer

```
;; APPLICANT: Steiner, Sabine
;; APPLICANT: Mohr, Christine
```

```
;; APPLICANT: wendland, jurgen
;; APPLICANT: knechtle, philipp
```

```

; APPLICANT: Rebischung, Corinne
; TITLE OF INVENTION: GENOMIC DNA

```

```

; TITLE OF INVENTION: AND USES THEREOF
;
; NUMBER OF SEQUENCES: 1152

```

; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: NO. 6239264artls Corporation

STREET: 3054 Cornwallis Road
CITY: Research Triangle Park

STATE: No. 6239264th Carolina
COUNTRY: USA

COMPUTER READABLE FORM:
ZIP: 27709

COMPUTER: IBM PC compatible
MEDIUM TYPE: Floppy disk

```

; CONSOLE: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
COEMWABP: Datoteka Policos #1 0   Verzies #1 30

```

```

; SOFTWARE:  falconlin release #1.0,
; CURRENT APPLICATION DATA:
; ADDITION NUMBER:  114/00/0000  110

```

APPLICATION NUMBER: US/0
FILING DATE: 24-DEC-1997

```

; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:

```

APPLICATION NUMBER: CH 0
FILING DATE: 31-DEC-1996

ATTORNEY/AGENT INFORMATION:
NAME: Meigs, J. Timothy

REGISTRATION NUMBER: 38,241
REFERENCE/DOCKET NUMBER: PF/5-30306/A/CGC1976

TELECOMMUNICATION INFORMATION
TELEPHONE: 919-541-8587

TELEFAX: 919-541-8689
INFORMATION FOR SEO ID NO: 547.

SEQUENCE CHARACTERISTICS:
LENGTH: 835 base pairs

TYPE: nucleic acid
STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE LIFE: DNA (g
ORIGINAL SOURCE:
ORIGIN: D001381ED

ORGANISM: PAG1381RP

US-08-998-416-547

| | | | | |
|-----------------------|--------------|----------------|-----------------|-------------|
| Query Match | 10.8%; | Score 38.8; | DB 4; | Length 835; |
| Best Local Similarity | 45.4%; | Pred. No. 2.3; | | |
| Matches 139; | Conservative | 0; | Mismatches 167; | Indels 0; |
| | | | | Gaps 0; |

QY 5 aaaaaagatgaaagtatttttattcttcctccttttttttgataaatcttaaatcatcttc 64

 | | | | | | | | | | | | | | | | | | | | | |

Db 122 AATAATGATTGTAGTTTAAATAATTAATAATTTGGTTATATCTAATAATTCCTTT 181

Qy 65 ttctgccaatgatatataa aaattcgcgataaacaataatattgcgatatcgttttttag 124
||| ||| ||| | ||||| || | ||
Db 182 ATTAGTAAATTTACTAATGAATTTAACTAATAATATATATAATTAATTTAGCTCTAA 2411

DY 125 ttcggtttgagaaaggattctcgacttctgaagtgcacgatataatgattggagc 186
|| ||| || | | | | ||| ||
DB 242 ttmgattatttaaacctcaagcaaatattttatctcatcttatatttaatttgacacacacat 307

Q7 185 taggttagtcttcgacattgtatgtgatgttctgtatattagtagtcgcacactta 244

245 aaccttaatgagcttctataagcccaattatataagatataacaagtgacaact 300

| | | | |
|----|-----|-------|-----|
| D6 | 302 | ttact | 310 |
| QY | 305 | ttact | 310 |

Db 422 TGTAGT 427

Search completed: October 17, 2001, 19:14:17
Job time: 15732 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 17, 2001, 16:25:13 ; Search time 5433.22 Seconds
(without alignments)
622.857 Million cell updates/sec

Title: US-09-502-426-1_COPY_6111_6468
Perfect score: 358
Sequence: 1 aaaaaaaaaagatgaagt.....tgtcaataccaacacag 358

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 10228115 seqs, 4726426750 residues

Total number of hits satisfying chosen parameters: 20456230

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

EST:*
1: gb_est1:*
2: gb_est2:*
3: gb_est3:*
4: gb_est4:*
5: gb_est5:*
6: gb_est6:*
7: gb_est7:*
8: gb_est8:*
9: gb_est9:*
10: gb_est10:*
11: gb_est11:*
12: gb_est12:*
13: gb_est13:*
14: gb_est14:*
15: gb_est15:*
16: gb_est16:*
17: gb_est17:*
18: gb_est18:*
19: gb_est19:*
20: gb_est20:*
21: gb_est21:*
22: gb_est22:*
23: gb_est23:*
24: gb_est24:*
25: gb_est25:*
26: gb_est26:*
27: gb_est27:*
28: gb_est28:*
29: gb_est29:*
30: gb_est30:*
31: gb_est31:*
32: gb_est32:*
33: gb_est33:*
34: gb_est34:*
35: gb_est35:*
36: gb_est36:*
37: gb_est37:*
38: gb_est38:*
39: gb_est39:*
40: gb_est40:*
41: gb_est41:*
42: gb_est42:*
43: gb_est43:*

44: em_estnum10:*
45: em_estnum11:*
46: em_estnum12:*
47: em_estnum13:*
48: em_estnum14:*
49: em_estnum15:*
50: em_estnum16:*
51: em_estnum17:*
52: em_estnum18:*
53: em_estnum19:*
54: em_estnum20:*
55: em_estnum21:*
56: em_estnum22:*
57: em_estnum23:*
58: em_estnum24:*
59: em_estnum25:*
60: em_estnum26:*
61: em_estnum27:*
62: em_estnum28:*
63: em_estnum29:*
64: em_estnum30:*
65: em_estnum31:*
66: em_estnum32:*
67: em_estnum33:*
68: em_estnum34:*
69: em_estnum35:*
70: em_estnum36:*
71: em_estnum37:*
72: em_estnum38:*
73: em_estnum39:*
74: em_estnum40:*
75: em_estnum41:*
76: em_estnum42:*
77: em_estnum43:*
78: em_estnum44:*
79: em_estnum45:*
80: em_estnum46:*
81: em_estnum47:*
82: em_estnum48:*
83: em_estnum49:*
84: em_estnum50:*
85: em_estnum51:*
86: em_estnum52:*
87: em_estnum53:*
88: em_estnum54:*
89: em_estnum55:*
90: em_estnum56:*
91: em_estnum57:*
92: em_estnum58:*
93: em_estnum59:*
94: em_estnum60:*
95: em_estnum61:*
96: em_estnum62:*
97: em_estnum63:*
98: em_estnum64:*
99: em_estnum65:*
100: em_estnum66:*
101: em_estnum67:*
102: em_estnum68:*
103: em_estnum69:*
104: em_estnum70:*
105: em_estnum71:*
106: em_estnum72:*
107: em_estnum73:*
108: em_estnum74:*
109: em_estnum75:*
110: em_estnum76:*
111: em_estnum77:*
112: em_estnum78:*
113: em_estnum79:*
114: em_estnum80:*
115: em_estnum81:*
116: em_estnum82:*

```

117: gb_est48:*
118: gb_est49:*
119: gb_est50:*
120: gb_est51:*
121: gb_est52:*
122: gb_est53:*
123: gb_est54:*
124: gb_est55:*
125: gb_est56:*
126: gb_est57:*
127: gb_est58:*
128: gb_est59:*
129: gb_est60:*
130: gb_est61:*
131: gb_est62:*
132: gb_est63:*
133: gb_est64:*
134: gb_est65:*
135: gb_est66:*
136: gb_est67:*
137: gb_est68:*
138: gb_est69:*
139: gb_est70:*
140: gb_est71:*
141: gb_est72:*
142: gb_est73:*
143: gb_est74:*
144: gb_est75:*
145: gb_est76:*
146: gb_est77:*
147: gb_est78:*
148: gb_est79:*
149: gb_est80:*
150: gb_est81:*
151: gb_est82:*
152: gb_est83:*
153: gb_est84:*
154: gb_est85:*
155: gb_est86:*
156: gb_est87:*
157: gb_est88:*
158: gb_est89:*
159: gb_est90:*
160: gb_est91:*
161: gb_est92:*
162: gb_est93:*
163: gb_est94:*
164: gb_est95:*
165: gb_est96:*
166: gb_est97:*
167: gb_est98:*
168: gb_est99:*
169: gb_est100:*
170: gb_est101:*
171: gb_est102:*
172: gb_est103:*
173: gb_est104:*
174: gb_est105:*
175: gb_est106:*
176: gb_est107:*
177: gb_est108:*
178: gb_est109:*
179: gb_est110:*
180: gb_est111:*
181: gb_est112:*
182: gb_est113:*
183: gb_est114:*
184: gb_est115:*
185: gb_est116:*
186: gb_est117:*
187: gb_est118:*
188: gb_est119:*
189: gb_est120:*

```

```

190: gb_est121:*
191: gb_est122:*
192: gb_est123:*
193: gb_est124:*
194: gb_est125:*
195: gb_est126:*
196: gb_est127:*
197: gb_est128:*
198: gb_est129:*
199: gb_est130:*
200: gb_est131:*
201: gb_est132:*
202: gb_est133:*
203: gb_est134:*
204: gb_est135:*
205: gb_est136:*
206: gb_est137:*
207: gb_est138:*
208: gb_est139:*
209: gb_est140:*
210: gb_est141:*
211: gb_est142:*
212: gb_est143:*
213: gb_est144:*
214: gb_est145:*
215: gb_est146:*
216: gb_est147:*
217: gb_est148:*
218: gb_est149:*
219: gb_est150:*
220: gb_est151:*
221: gb_est152:*
222: gb_est153:*
223: gb_est154:*
224: gb_est155:*
225: gb_est156:*
226: gb_est157:*
227: gb_est158:*
228: gb_est159:*
229: gb_est160:*
230: gb_est161:*
231: gb_est162:*
232: gb_est163:*
233: gb_est164:*
234: gb_est165:*
235: gb_est166:*
236: gb_est167:*
237: gb_est168:*
238: gb_est169:*
239: gb_est170:*
240: gb_est171:*
241: gb_est172:*
242: gb_est173:*
243: gb_est174:*
244: gb_est175:*
245: gb_est176:*
246: gb_est177:*
247: gb_est178:*
248: gb_est179:*
249: gb_est180:*
250: gb_est181:*
251: gb_est182:*
252: gb_est183:*
253: gb_est184:*
254: gb_est185:*
255: gb_est186:*
256: gb_est187:*
257: gb_est188:*
258: gb_est189:*

```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

| | |
|-----------------------|--|
| JOURNAL | Submitted (11-JUN-1999) Genoscope - Centre National de Sequencage BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr - Web : www.genoscope.cns.fr) |
| COMMENT | Determination of this BAC-end sequence was carried out as part of a collaboration with the Berkeley Drosophila Genome Project (BDGP). The BDGP is constructing a physical map of the Drosophila melanogaster genome using these BACS. For further information please see: http://www.fruitfly.org/The BDGP Drosophila melanogaster BAC library was prepared by Kazutoyo Osoegawa and Aaron Mammeter in Pieter de Jong's laboratory in the Department of Cancer Genetics at the Roswell Park Cancer Institute in Buffalo, NY. The library is named RPCI-98 and was constructed by partial EcoRI digestion of Drosophila DNA provided by the BDGP from the isogenic strain y2; cn bw sp. The same strain used for the BDGP's p1 and EST libraries. A more detailed description of the library and how to order individual BAC clones, the entire library, or filters for hybridization from the BACPAC Resource Center can be found at http://bacpac.med.buffalo.edu/drosophila_bac.htm . |
| FEATURES | location/Qualifiers 1..1101 /organism="Drosophila melanogaster" /db_xref="taxon:7227" /clone_lib="RPCI-98" /clone="BACR48P19" /note="end : TET3" |
| BASE COUNT | 469 a 6 c 69 g 151 t 406 others |
| ORIGIN | |
| Query Match | 19.0%; Score 68; DB 219; Length 1101; |
| Best Local Similarity | 26.4%; Pred. No. 0.0013; |
| Matches | 68; Conservative 100; Mismatches 90; Indels 0; Gaps 0; |
| QY | 4 aaaaaagaagaaagtatttlaatccttcctcttttttttgataaatcaatcatl 63 |
| Db | 603 WAKKMWADFAATTTTWTTTTTTTTMMATTTTTTTMTAAATTTTTTTTATAKT 662 |
| QY | 64 ttltggccaatgatataaaaatttgataaataatatattgatatatcgltttta 123 |
| Db | 663 TTTTCTTTTATTAAATAAANAAMWTTDTMAAAMWTTTKKKKKKAADKWKDAKKMDGA 722 |
| QY | 124 gtctgggttggaagaagggttgcacttcgtgaagagtgacgatagatatggag 183 |
| Db | 723 KKARTTKKDKKAAWAADKDRKKGGKKGKGGKGGKGGKGGKGGKGGKGGKGGK 782 |
| QY | 184 cttagtgttagcttggaacatttgatggatgttgtgatattagtgtagcacatt 243 |
| Db | 783 AGDDAKKXTKKKKKAATTTTKKKKKGKKKKKAARKKADRTKTXWDAAAAAAAANK 842 |
| QY | 244 aaacctaaatcggttt 261 |
| Db | 843 TKDKGKKKKKKKKTKTK 860 |
| RESULT 2 | |
| CNS00DKY | 928 bp DNA GSS 04-JUN-1999 |
| LOCUS | Drosophila melanogaster genome survey sequence T7 end of BAC # |
| DEFINITION | BACR27A24 of RPCI-98 library from Drosophila melanogaster (fruit fly), genomic survey sequence. |
| ACCESSION | AL071865 |
| VERSION | AL071865.1 GI:4948170 |
| KEYWORDS | GSS. |
| SOURCE | fruit fly. |
| ORGANISM | Drosophila melanogaster |
| REFERENCE | Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila. |
| AUTHORS | Genoscope. Direct Submission |
| TITLE | Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage |
| JOURNAL | BP 19191006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr |

otidae; Tetraodon.

High quality sequence start: 48
High quality sequence stop: 291.
Location/Qualifiers
1. .961
/organism="Entamoeba histolytica"
/strain="HMI:IMMS"
/db_xref="taxon:5759"
/clone_lib="Entamoeba histolytica Sheared DNA"
/note="Vector: PHOS1; Site: 1; Bst I; Constructed at The Institute for Genomic Research (TIGR), Rockville, MD. Genomic DNA isolated from broth cultures of E. histolytica using a method described by Clark and Diamond (Clark, C.G., and Diamond, L.S. (1993) Entamoeba histolytica: a method for isolate identification. Exp. Parasitol. 77:450.). The DNA was mechanically sheared to give a light size distribution (~2 kb). The v + i method used for the library construction is described in detail in Smith, H.O. and Venter, J.C. (Making small insert libraries for whole genome shotgun sequencing projects. In Genome Sequencing: A Practical Approach, eds. M. Vaudin and B. Barrell, Oxford University Press, 1999)."

BASE COUNT 518 a 63 c 91 g 289 t
ORIGIN

Query Match 16.1%; Score 57.8; DB 248; Length 961;
Best Local Similarity 64.7%; Pred. No. 0.11;
Matches 86; Conservative 0; Mismatches 47; Indels 0; Gaps 0;

QY 1 aaaaaaaaaaagatgaagatatttattctctctctttttttgataatttaacaa 60
Db 674 AAAAAAAAAAATTTTATTTATTTATTTATTTATTTATTTATTTATTTATTTA 615
QY 61 ttttttgcacatgatataataatttgataataataatattgagatcgattt 120
Db 614 TTTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTT 555
QY 121 ttagtcgagttt 133
Db 554 TTTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTT 542

RESULT 7
CNS00EJ4/c 1101 bp DNA GSS 04-JUN-1999
LOCUS Drosophila melanogaster genome survey sequence T7 end of BAC:
DEFINITION BACR29K22 of RPCI-98 library from Drosophila melanogaster (fruit fly), genomic survey sequence.
ACCESSION AL069257.1 GI:4949400
VERSION AL069257
KEYWORDS GSS.
SOURCE fruit fly.
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachyera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE 1 (bases 1 to 1101)
Genoscope.
Direct Submission
Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
- Web : www.genoscope.cns.fr)
Determination of this BAC-end sequence was carried out as part of a collaboration with the Berkeley Drosophila Genome Project (BDGP).
The BDGP is constructing a physical map of the Drosophila melanogaster genome using these BACs. For further information please see <http://www.fruitfly.org> The BDGP Drosophila melanogaster BAC library was prepared by Kazutoyo Osoegawa and Aaron Mamoser in Pieter de Jong's laboratory in the Department of Cancer Genetics at the Roswell Park Cancer Institute in Buffalo, NY. The library is named RPCI-98 and was constructed by partial EcoRI digestion of Drosophila DNA provided by the BDGP from the isogenic strain y2; cn bw sp, the same strain used for the BDGP's

P1 and EST libraries. A more detailed description of the library and how to order individual BAC clones, the entire library, or filters for hybridization from the BACAC Resource Center can be found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.
Location/Qualifiers
1. .1101
/organism="Drosophila melanogaster"
/db_xref="taxon:7227"
/clone_lib="RPCI-98"
/note="BACR29K22"
/note="end : T7"

BASE COUNT 434 a 140 c 25 g 377 t 125 others
ORIGIN

Query Match 15.9%; Score 57; DB 219; Length 1101;
Best Local Similarity 42.1%; Pred. No. 0.15; Mismatches 163; Indels 7; Gaps 1;
Matches 149; Conservative 35

QY 1 aaaaaaaaaaagatgaagatatttattctctctctttttttgataatttaacaa 60
Db 895 AAAAAAAAAAATTTTAAATTTTATTTTAAATTTTATTTTAAATTTTAAATTTTAA 836
QY 61 ttttttgcacatgatataataatttgataataataatattgagatcgattt 120
Db 835 TAAATTTAAAD 776
QY 121 ttagtcgagttt-----gagaaaggttcgacttcgaaagcgagatgat 173
Db 775 TTTTATTAATGAGTATTTTGGGAGTAAAGGGTAAATTTTATTAATTTATTTATTT 716
QY 174 agattggagcagtcagtcgacttcgacattgatgatgagtcgtgatattatgt 233
Db 715 TTTTATTAATGAGTATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTT 656
QY 234 cgacactataaaccttaaatggccttcataaagcccaatataacgataacaa 293
Db 655 TTTTAAATTAATTTAAATTTAAATTTAAATTTAAATTTAAATTTAAATTTAA 596
QY 294 aagtgacacttacttcgcttcgttcgacgagcaataaacaattgcaata 347
Db 595 ATGNNNTGKGTGTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTTGTG 542

RESULT 8
CNS018B6/c 894 bp DNA GSS 26-JUL-1999
LOCUS Drosophila melanogaster genome survey sequence T7 end of BAC:
DEFINITION BACN13B16 of DrosBAC library from Drosophila melanogaster (fruit fly), genomic survey sequence.
ACCESSION AL109126
VERSION AL109126
KEYWORDS GSS.
SOURCE fruit fly.
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachyera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE 1 (bases 1 to 894)
Genoscope.
Direct Submission
Submitted (23-JUL-1999) Genoscope - Centre National de Sequencage :
BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
- Web : www.genoscope.cns.fr)
Determination of this BAC-end sequence was carried out as part of a collaboration with the European Drosophila Genome Project (EDGP) - <http://www.edgp.ebi.ac.uk> - This Drosophila melanogaster BAC library (Dros BAC) was made by Alain Billard at CEPH (Centre d'Etude du Polymorphisme Humain) with funding provided by a MRC project grant. The DNA was prepared from embryos by Alain Bucheton and Genevieve Payan. It has been constructed from the vector pBelobAC11.

COMMENT
AUTHORS
TITLE
JOURNAL
FEATURES
Location/Qualifiers

[illegible]

| | | | | | |
|-----------|------------|---|------------|-----|-------------|
| | RESULT | 11 | | | |
| | Locus | CNS04AIH/C | | | |
| | DEFINITION | CNS04AIH | 949 bp | DNA | GSS |
| | | Tetraodon nigroviridis genome survey sequence T7 end of clone | | | 21-MAY-2000 |
| | | 095F19 of library G from Tetraodon nigroviridis, genomic survey | | | |
| | | sequence. | | | |
| | ACCESSION | AL281906 | | | |
| | VERSION | AL281906.1 | GI:8020236 | | |
| | KEYWORDS | GSS; genome survey sequence. | | | |
| | SOURCE | Tetraodon nigroviridis.. | | | |
| | ORGANISM | Tetraodon nigroviridis | | | |
| | | Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; | | | |
| | | Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei; | | | |
| | | Acanthomorpha; Acanthopterygii; Perciformes; Tetraodontiformes; | | | |
| | | Tetraodontidae; Tetraodon. | | | |
| REFERENCE | | 1 (bases 1 to 949) | | | |
| AUTHORS | | Roest-Crolius,H., Jalllon,O., Dasilva,C., Fizames,C., Fisher,C., | | | |
| | | Bouneau,L., Billault,A., Quetler,F., Saurin,W., Bernot,A. and | | | |
| | | Weissenbach,J. | | | |
| TITLE | | Characterization and repeat analysis of the compact genome of the | | | |
| JOURNAL | | Freshwater pufferfish Tetraodon nigroviridis | | | |
| REFERENCE | | Unpublished | | | |
| AUTHORS | | 2 (bases 1 to 949) | | | |
| | | Roest-Crolius,H., Jalllon,O., Dasilva,C., Bouneau,L., Fisher,C., | | | |
| | | Bernot,A., Fizames,C., Wincker,P., Brottier,P., Quetler,F., | | | |
| | | Saurin,W. and Weissenbach,J. | | | |
| TITLE | | Human gene number estimate provided by genome wide analysis using | | | |
| JOURNAL | | Tetraodon nigroviridis DNA sequence | | | |
| REFERENCE | | Unpublished | | | |
| AUTHORS | | 3 (bases 1 to 949) | | | |
| | | Genoscope. | | | |
| TITLE | | Direct Submission | | | |
| JOURNAL | | Submitted (12-APR-2000) to the EMBL/Genbank/DBJ databases | | | |
| COMMENT | | This sequence is a single read and was generated as part of a large | | | |
| | | scale clone-and-sequencing project of the Tetraodon nigroviridis | | | |
| | | genome. For more information, please take a look at | | | |
| | | http://www.genoscope.cns.fr/Tetraodon . | | | |

```

FEATURES
source
    location/Qualifiers
    1..949
    /organism="Tetrarodon nigroviridis"
    /db_xref="taxon:99883"
    /clone="095F19"
    /clone_11b="G"
    /note="Genoscope sequence ID : C0BG095CC10LPI-end : T7"
BASE COUNT
355 a 144 c 185 g 223 t 42 others
ORIGIN

```

| | | | | |
|-----------------------|-------|-----------------|---------|----------------|
| Query Match | 15.6% | Score 56; | DB 221; | Length 949; |
| Best Local Similarity | 49.3% | Pred. NO. 0.23; | | |
| Matches | 74; | Conservative | 21; | Mismatches 55; |
| | | | Indels | 0; |
| | | | Gaps | 0; |

[illegible]

| | |
|-------------|--|
| RESULT | 12 |
| CNS000AQX/c | |
| LOCUS | CNS000AQX 767 bp DNA GSS 03-JUN-1999 |
| DEFINITION | Drosophila melanogaster genome survey sequence TET3 end of BAC # |
| | BACR21A20 of RPi1-98 library from Drosophila melanogaster (fruit |
| | fly), genomic survey sequence. |
| ACCESSION | AL055924 |
| VERSION | AL055924.1 GI:4936693 |
| KEYWORDS | |
| SOURCE | GS. |
| ORGANISM | fruit fly, Drosophila melanogaster |
| | Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta; |
| | Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; |
| | Muscomorpha; Ephydroidea; Drosophilidae; Drosophila. |
| REFERENCE | 1 (bases 1 to 767) |

JOURNAL Submitted (02-JUN-1999) Genoscope - Centre National de Séquençage
BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
COMMENT - Web : www.genoscope.cns.fr)
Determination of this BAC-end sequence was carried out as part of a
collaboration with the Berkeley Drosophila Genome Project (BDGP).
The BDGP is constructing a physical map of the Drosophila
melanogaster genome using these BACs. For further information
please see <http://www.fruitfly.org/TheBDGP/Drosophila>
melanogaster BAC library was prepared by Kazutoyo Oosawa and
Aaron Mamoser in Pieter de Jong's laboratory in the Department of
Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
NY. The library is named RPCI-98 and was constructed by partial
EcoRI digestion of Drosophila DNA provided by the BDGP from the
isogenic strain y2: cn bw sp. the same strain used for the BDGP's
p1 and EST libraries. A more detailed description of the library
and how to order individual BAC clones, the entire library, or
filters for hybridization from the BACPAC Resource center can be
found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.

```
FEATURES
source      location/Qualifiers
            1. .767
/organism="Drosophila melanogaster"
/db_xref="taxon:7227"
/clone_lib="RPIC1-98"
/clone="BACR21A20"
/note="end : TERC"

BASE COUNT      312 a      105 c      65 g      164 t      121 others
ORIGIN
```

| | | | | |
|-----------------------|--------|-----------------|---------|-----------------|
| Query Match | 15.6%; | Score 55.8; | DB 219; | Length 767; |
| Best Local Similarity | 33.1%; | Pred. NO. 0.25; | | |
| Matches | 99; | Conservative | 64; | Mismatches 136; |
| | | | Indels | 0; |
| | | | Gaps | 0; |

| | | | |
|----|-----|---|-----|
| Oy | 23 | tttatactctctctctcttttcttbaataatctbaatcattcttgcacaaagataat | 82 |
| | | : : : : : : : : : | |
| Db | 389 | TATTMTMTTTTTTTTTTGTTTTTMMTWMTCKMTYMTKMTTMTTCTTKTKTKMKKTTT | 300 |
| Oy | 83 | aaaaatttgataataatattatgatgatcgttttaagtcggggtctggaaagg | 142 |
| | | : : : : : : : : | |
| Db | 329 | TTTGTGTAATKATKMMMTTAKMKMTTTTTTTTTTTTTTTTTTTTMMATMAGAGMAAK | 270 |
| Oy | 143 | gttcgcacttcgaagtcgaagatgatataagatcggagcctaggttcgcttcggac | 202 |

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 17, 2001, 19:14:21 ; Search time 44.67 Seconds
(without alignments)
696.219 Million cell updates/sec

Title: US-09-502-426-2
Perfect score: 2681
Sequence: 1 MFETEHHTLLPLLLPLSLLS.....FAFPVDFPNCGLPIRVSRLL 513

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 412676 seqs, 60623988 residues
Total number of hits satisfying chosen parameters: 412676

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database : A_Geneseq_0601.*
1: /SIDSL/gcgdata/geneseq/geneseqp/AA1980.DAT:*
2: /SIDSL/gcgdata/geneseq/geneseqp/AA1981.DAT:*
3: /SIDSL/gcgdata/geneseq/geneseqp/AA1982.DAT:*
4: /SIDSL/gcgdata/geneseq/geneseqp/AA1983.DAT:*
5: /SIDSL/gcgdata/geneseq/geneseqp/AA1984.DAT:*
6: /SIDSL/gcgdata/geneseq/geneseqp/AA1985.DAT:*
7: /SIDSL/gcgdata/geneseq/geneseqp/AA1986.DAT:*
8: /SIDSL/gcgdata/geneseq/geneseqp/AA1987.DAT:*
9: /SIDSL/gcgdata/geneseq/geneseqp/AA1988.DAT:*
10: /SIDSL/gcgdata/geneseq/geneseqp/AA1989.DAT:*
11: /SIDSL/gcgdata/geneseq/geneseqp/AA1990.DAT:*
12: /SIDSL/gcgdata/geneseq/geneseqp/AA1991.DAT:*
13: /SIDSL/gcgdata/geneseq/geneseqp/AA1992.DAT:*
14: /SIDSL/gcgdata/geneseq/geneseqp/AA1993.DAT:*
15: /SIDSL/gcgdata/geneseq/geneseqp/AA1994.DAT:*
16: /SIDSL/gcgdata/geneseq/geneseqp/AA1995.DAT:*
17: /SIDSL/gcgdata/geneseq/geneseqp/AA1996.DAT:*
18: /SIDSL/gcgdata/geneseq/geneseqp/AA1997.DAT:*
19: /SIDSL/gcgdata/geneseq/geneseqp/AA1998.DAT:*
20: /SIDSL/gcgdata/geneseq/geneseqp/AA1999.DAT:*
21: /SIDSL/gcgdata/geneseq/geneseqp/AA2000.DAT:*
22: /SIDSL/gcgdata/geneseq/geneseqp/AA2001.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description |
|------------|-------|-------------|--------|----|-----------------------------|
| 1 | 2681 | 100.0 | 513 | 21 | AA807921 A cytochrome P450 |
| 2 | 978.5 | 36.5 | 472 | 18 | AAW27153 Arabidopsis thalia |
| 3 | 978.5 | 36.5 | 472 | 21 | AA844571 Arabidopsis thalia |
| 4 | 978.5 | 36.5 | 472 | 21 | AA845022 Arabidopsis thalia |
| 5 | 978.5 | 36.5 | 491 | 21 | AA845021 Arabidopsis thalia |
| 6 | 978.5 | 36.5 | 492 | 21 | AA844570 Arabidopsis thalia |
| 7 | 946 | 35.3 | 444 | 21 | AA844572 Arabidopsis thalia |
| 8 | 946 | 35.3 | 444 | 21 | AA845023 Arabidopsis thalia |
| 9 | 726 | 27.1 | 461 | 21 | AA846490 Arabidopsis thalia |
| 10 | 726 | 27.1 | 462 | 21 | AA846489 Arabidopsis thalia |
| 11 | 726 | 27.1 | 465 | 21 | AA846491 Arabidopsis thalia |

| Result No. | Score | Query Match | Length | ID | Description |
|------------|-------|-------------|--------|----|------------------------------|
| 12 | 712 | 26.6 | 461 | 21 | AA811836 Arabidopsis thalia |
| 13 | 712 | 26.6 | 462 | 21 | AA811835 Arabidopsis thalia |
| 14 | 712 | 26.6 | 465 | 21 | AA811834 Arabidopsis thalia |
| 15 | 653.5 | 24.4 | 468 | 21 | AA830783 Arabidopsis thalia |
| 16 | 606 | 22.6 | 471 | 21 | AA833013 Arabidopsis thalia |
| 17 | 606 | 22.6 | 479 | 21 | AA833012 Arabidopsis thalia |
| 18 | 603 | 22.3 | 469 | 21 | AA833014 Arabidopsis thalia |
| 19 | 598.5 | 22.3 | 430 | 21 | AA820784 Arabidopsis thalia |
| 20 | 547.5 | 20.4 | 489 | 21 | AA830048 Arabidopsis thalia |
| 21 | 546 | 20.4 | 481 | 21 | AA830049 Arabidopsis thalia |
| 22 | 528.5 | 19.7 | 388 | 21 | AA820785 Arabidopsis thalia |
| 23 | 503 | 18.8 | 492 | 19 | AAW37733 Arabidopsis thalia |
| 24 | 503 | 18.8 | 492 | 19 | AAW44159 Arabidopsis thalia |
| 25 | 490 | 18.3 | 433 | 21 | AA830050 Arabidopsis thalia |
| 26 | 461 | 17.2 | 497 | 19 | AAW37735 Arabidopsis thalia |
| 27 | 461 | 17.2 | 497 | 19 | AAW44161 Arabidopsis thalia |
| 28 | 439 | 16.4 | 497 | 19 | AAW37734 Arabidopsis thalia |
| 29 | 439 | 16.4 | 497 | 19 | AAW44160 Arabidopsis thalia |
| 30 | 420 | 15.7 | 253 | 21 | AA817903 Arabidopsis thalia |
| 31 | 411.5 | 15.3 | 237 | 21 | AA817905 Arabidopsis thalia |
| 32 | 411.5 | 15.3 | 242 | 21 | AA817904 Arabidopsis thalia |
| 33 | 319 | 11.9 | 519 | 21 | AA842953 Arabidopsis thalia |
| 34 | 318 | 11.9 | 519 | 21 | AA816752 Arabidopsis thalia |
| 35 | 316 | 11.8 | 500 | 21 | AA842954 Arabidopsis thalia |
| 36 | 315 | 11.7 | 500 | 21 | AA816753 Arabidopsis thalia |
| 37 | 298.5 | 11.1 | 905 | 21 | AA838596 Arabidopsis thalia |
| 38 | 298.5 | 11.1 | 949 | 21 | AA838595 Arabidopsis thalia |
| 39 | 298.5 | 11.1 | 1041 | 21 | AA838594 Arabidopsis thalia |
| 40 | 297 | 11.1 | 501 | 21 | AA850190 Arabidopsis thalia |
| 41 | 297 | 11.1 | 516 | 21 | AA850189 Arabidopsis thalia |
| 42 | 290.5 | 10.8 | 502 | 21 | AA829761 Arabidopsis thalia |
| 43 | 289 | 10.8 | 511 | 20 | AA833341 Arabidopsis thalia |
| 44 | 289 | 10.8 | 511 | 21 | AA819694 Arabidopsis thalia |
| 45 | 285.5 | 10.6 | 502 | 21 | AA8191348 Arabidopsis thalia |

ALIGNMENTS

RESULT 1
AAB07921 standard; Protein: 513 AA.
ID AAB07921;
AC AAB07921;
DT 14-NOV-2000 (first entry)
DE A cytochrome P450 enzyme designated DWf4.
KW DWf4; cytochrome P450 enzyme; brassinosteroid; 22alpha-hydroxylation;
KW plant phenotype; cell elongation.
XX Arabidopsis sp.
XX OS
XX PN WO200047715-A2.
XX PD 17-AUG-2000.
XX PF 11-FEB-2000; 2000WO-US03820.
XX PR 11-FEB-1999; 99US-0119657.
XX PR 11-FEB-1999; 99US-0119658.
XX PA (ARIZ-) ARIZONA BOARD OF REGENTS.
XX PI Azpiroz R, Choe S, Feldmann KA;
XX WPI: 2000-549142/50.
XX N-PSDB: AAA59599.
XX PT New isolated dwf4 polynucleotide useful for altering the phenotype of
XX plants, for diagnostic assays and in the production of antibodies -

OY 491 DQPPFPVDFPNCGLPIRVSR 511
|: || || |
Db 447 dklvfpftrtqrkryplfvkr 467

RESULT 3

AAC44571

ID AAG44571 standard; Protein: 472 AA.

AC AAG44571;

DT 18-OCT-2000 (first entry)

DE Arabidopsis thaliana protein fragment SEQ ID NO: 55847.

KW Protein identification: signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.

XX Arabidopsis thaliana.

XX EPI033405-A2.

XX PD 06-SEP-2000.

XX PF 25-FEB-2000; 2000EP-0301439.

XX PR 25-FEB-1999; 99US-0121825.

XX PR 05-MAR-1999; 99US-0123180.

XX PR 09-MAR-1999; 99US-0123548.

XX PR 23-MAR-1999; 99US-0125788.

XX PR 25-MAR-1999; 99US-0126264.

XX PR 29-MAR-1999; 99US-0126785.

XX PR 01-APR-1999; 99US-0127452.

XX PR 06-APR-1999; 99US-0128234.

XX PR 08-APR-1999; 99US-0128714.

XX PR 16-APR-1999; 99US-0129845.

XX PR 19-APR-1999; 99US-0130077.

XX PR 21-APR-1999; 99US-0130449.

XX PR 23-APR-1999; 99US-0130510.

XX PR 23-APR-1999; 99US-0130891.

XX PR 28-APR-1999; 99US-0131449.

XX PR 30-APR-1999; 99US-0132048.

XX PR 30-APR-1999; 99US-0132407.

XX PR 04-MAY-1999; 99US-0132484.

XX PR 05-MAY-1999; 99US-0132485.

XX PR 06-MAY-1999; 99US-0132486.

XX PR 06-MAY-1999; 99US-0132487.

XX PR 07-MAY-1999; 99US-0132863.

XX PR 11-MAY-1999; 99US-0134256.

XX PR 14-MAY-1999; 99US-0134218.

XX PR 14-MAY-1999; 99US-0134219.

XX PR 14-MAY-1999; 99US-0134221.

XX PR 14-MAY-1999; 99US-0134370.

XX PR 18-MAY-1999; 99US-0134768.

XX PR 19-MAY-1999; 99US-0134941.

XX PR 20-MAY-1999; 99US-0135124.

XX PR 21-MAY-1999; 99US-0135353.

XX PR 24-MAY-1999; 99US-0135629.

XX PR 25-MAY-1999; 99US-0136021.

XX PR 27-MAY-1999; 99US-0136392.

XX PR 28-MAY-1999; 99US-0136782.

XX PR 01-JUN-1999; 99US-0137222.

XX PR 03-JUN-1999; 99US-0137528.

XX PR 04-JUN-1999; 99US-0137502.

XX PR 07-JUN-1999; 99US-0137724.

XX PR 08-JUN-1999; 99US-0138094.

XX PR 10-JUN-1999; 99US-0138540.

XX PR 10-JUN-1999; 99US-0138847.

XX PR 14-JUN-1999; 99US-0139119.

XX PR 16-JUN-1999; 99US-0139452.

XX PR 16-JUN-1999; 99US-0139453.

XX PR 17-JUN-1999; 99US-0139492.

PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.

PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152353.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161982.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 36.5% Score 978.5; DB 21; Length 472;
Best Local Similarity 41.9%; Pred. No. 3e-78; Indels 41; Gaps 8;
Matches 210; Conservative 79; Mismatches 171;

QY 12 LLLPSTLLFLILKRRNRKTRFNLPGKSGMPFLGFTIGYIKPYTATTLDGFMQOH 71
DB 7 LLLLSlaagfl--lllrtrrmglppslpllgelrfqllgykutenpepflder 63
QY 72 VSKYKGRIRSLWEGEPTIVSADAGLNRPIILONBGRLEFCSYPRISIGILGKWSMLVLVD 131
DB 64 varygsvfmlthfgeptlfdsadpetnrfvlgneqklfecsyasincilgshsillmkgs 123

QY 132 MHRDRSISLNFSLHARLRTILKDKVERHTLEVLDSMOONSIFSADBEAKKFTFNLMAKH 191
DB 124 ltkrtmbsltmsfansslikdhlmdidltvfnldsswryll--meaakktlelvkq 181
QY 192 TMSMDPGEETEOULKEVEYTEMKGVSAPLNPGTAYHKALOSRAITLKFIERKMEERKL 251
DB 182 lmsfdpg-eweselsrkeylllviegffsfpilplfstlykqldar-----rkvaelt 232
QY 252 DIKEEDOEFEFVKTEDEKEMSKSDHVRKQRTDDLLGKSNVLKHSNLSTFOILLLSLFLA 311
DB 233 vvmkrtreeeeegae-----rkkdmlaallaaadgfsdeelvdfivaliva 278
QY 312 GHETSSVAIALAIFFLQACPKAVEELREHLEIRAKKELEESLNMDDYKKMPTOCVI 371
DB 279 gyetstlmtlavflicetplalaglkheekiramksd--syelswyskmpftcgv 336
QY 372 NETLRIGNVRFELHRAKALKDVRKGYDIPSGKVLVYISAVHLDNSRYDQNLFPWRMQ 431
DB 337 netlrvanligvfrfamtdeivkgykirkqwkvfssfravhldpnhtkdarftfnprwq 396
QY 432 QONNGASSSGSSTWGN-N-YMFFGGGPRLCAGSELAKLEMAVFIIHLVLFKFWELAED 490
DB 397 -----snvltgspnvftfpf999prlcp9elatarvalsvflnrlvqfswpaeg 446
QY 491 DQPAFPFVDFPNGLPIRVSF 511
DB 447 dkivfptrtqrkryplfvkr 467

RESULT 4
AAG45022
ID AAG45022 standard; Protein: 472 AA.
AC AAG45022;
XX 18-OCT-2000 (first entry)
DT
XX
DE Arabidopsis thaliana protein fragment SEQ ID NO: 56469.
XX
KW Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.
XX
OS Arabidopsis thaliana.
PN EP1033405-A2.
XX
XX 06-SEP-2000.
PD
XX
PF 25-FEB-2000; 2000EP-0301439.
XX
PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.

PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140699.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144332.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.

PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147260.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 30-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151308.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157153.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.

PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 36.5% Score 978.5; DB 21; Length 472;
Best Local Similarity 41.9%; Pred. No. 3e-78;
Matches 210; Conservative 79; Mismatches 171; Indels 41; Gaps 8;

QY 12 LLLPSLTLFLILKRRNRKTRFNLPGKSGWPLGFTTGYLKYATATLGDPMOQH 71
DQ 7 LLLSLTAAGFL---LIRTRIRYRMGLPPSLGLPIGLGETFGIAGKtenpepfider 63
QY 72 VSKGKIRKSLFEGPTVVSADAGINRPILOEGRLFCSTPRSTIGILGKWSMLVLVD 131
DQ 64 VARYSVMTLHfgepltfdsadpelnrfvlqneqklfecsyasicnllgkhslllmkgs 123
QY 132 MHRDRSISLNLFSHARLRTILKOVERHTLFLVDSMQONSIFSAODRAKKEFTNLMAKH 191
DQ 124 Lhkrmhsilmsfmsaiikhdlmldidrlvtrfnldswsrrvll--meaakktfelvkg 181
QY 192 IMSMDPGEETEOALKREYVTFMGVSNAPNLPGTAHKALOSRAITILKFERKMEERKL 251
DQ 182 lmsfdpg-ewesalrkeyllvlegffslplrfstfykaigar-----rkvaalt 232
QY 252 DIKEEDOEDEEVKEDKEMASKSDHVRKQRTDDLLGVNLKNSLSTQIIDLILSLFA 311
DQ 233 vvmkrrteeeegae-----rkkdmllaadagfsdeedvdfivaliva 278
QY 312 GHETSSVALAIFELQCPKAVEELREHELEIARAKKEGSEELNMDYKKMPTOCVI 371
DQ 279 gyetstslmtlavflctcpialaqklkeehkirkamsd--syslewsdyksmpftgcvv 336
QY 372 NETRLGVNVPFLRRKALKDVRKGYDIPSGKYLPLVISAHLNDSRYDQNLFPNWMQ 431
DQ 337 netlrvalilgvtfrantdvclkykikipkgwkvfssfravhldpnhkdaartlmpwrq 396
QY 432 QONNGASSSGSFTWGN--YMPGCGPRLCAGSESLAKLEMAVFIHHLVLFKFWMLAED 490
DQ 397 -----snavltgpnvfcprfggprlcpgyelarvalsvflhrlvlgfswvpaq 446
QY 491 DQPAFPVDFPNCPLRVSF 511
DQ 447 dklvffprrtqrkypflvkr 467

RESULT 5
AAG45021
ID AAG45021 standard; Protein: 491 AA.

XX AAG45021;

XX 18-OCT-2000 (first entry)

XX Arabidopsis thaliana protein fragment SEQ ID NO: 56468.

KW Protein identification: signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.

XX Arabidopsis thaliana.

XX EPI033405-A2.

XX

PD 06-SEP-2000.
XX 25-FEB-2000; 2000EP-0301439.
PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139453.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.

Db 416 -----snsvltspnvtfpfg99prrlcpqyelarvalsvflhrhltcfsfwpaed 465
Oy 491 DOPFAFPVDFPNCGRIVSR 511
| | | | |
Db 466 dklvfpctrcqkryplvkr 486

RESULT 6
AAG44570
ID AAG44570 standard; Protein: 492 AA.

XX AAG44570;
XX
XX 18-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana protein fragment SEQ ID NO: 55846.
XX
KW Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.

XX Arabidopsis thaliana.
XX
XX EP1033405-A2.
XX
PD 06-SEP-2000.
XX
XX
PE 25-FEB-2000; 2000EP-0301439.
XX
XX 25-FEB-1999; 99US-0121825.
XX 05-MAR-1999; 99US-0123180.
XX 09-MAR-1999; 99US-0123548.
XX 23-MAR-1999; 99US-0125788.
XX 25-MAR-1999; 99US-0126264.
XX 29-MAR-1999; 99US-0126785.
XX 01-APR-1999; 99US-0127462.
XX 06-APR-1999; 99US-0128234.
XX 08-APR-1999; 99US-0128714.
XX 16-APR-1999; 99US-0129845.
XX 19-APR-1999; 99US-0130077.
XX 21-APR-1999; 99US-0130449.
XX 23-APR-1999; 99US-0130510.
XX 28-APR-1999; 99US-0130891.
XX 30-APR-1999; 99US-0131449.
XX 05-MAY-1999; 99US-0132048.
XX 06-MAY-1999; 99US-0132485.
XX 07-MAY-1999; 99US-0132486.
XX 11-MAY-1999; 99US-0132863.
XX 14-MAY-1999; 99US-0134218.
XX 14-MAY-1999; 99US-0134219.
XX 14-MAY-1999; 99US-0134221.
XX 18-MAY-1999; 99US-0134370.
XX 19-MAY-1999; 99US-0134768.
XX 20-MAY-1999; 99US-0134941.
XX 21-MAY-1999; 99US-0135124.
XX 21-MAY-1999; 99US-0135353.
XX 24-MAY-1999; 99US-0135629.
XX 25-MAY-1999; 99US-0136021.
XX 27-MAY-1999; 99US-0136392.
XX 28-MAY-1999; 99US-0136782.
XX 01-JUN-1999; 99US-0137222.
XX 03-JUN-1999; 99US-0137528.
XX 04-JUN-1999; 99US-0137502.
XX 07-JUN-1999; 99US-0137724.
XX 08-JUN-1999; 99US-0138094.
XX 10-JUN-1999; 99US-0138540.
XX 10-JUN-1999; 99US-0138847.
XX 14-JUN-1999; 99US-0139119.
XX 16-JUN-1999; 99US-0139452.

PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139850.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.

PR 13-AUG-1999; 99US-0148565;
PR 13-AUG-1999; 99US-0148684;
PR 16-AUG-1999; 99US-0149368;
PR 17-AUG-1999; 99US-0149175;
PR 18-AUG-1999; 99US-0149426;
PR 20-AUG-1999; 99US-0149722;
PR 20-AUG-1999; 99US-0149723;
PR 20-AUG-1999; 99US-0149929;
PR 23-AUG-1999; 99US-0149902;
PR 23-AUG-1999; 99US-0149930;
PR 25-AUG-1999; 99US-0150366;
PR 26-AUG-1999; 99US-0150884;
PR 27-AUG-1999; 99US-0151065;
PR 27-AUG-1999; 99US-0151066;
PR 27-AUG-1999; 99US-0151080;
PR 30-AUG-1999; 99US-0151303;
PR 31-AUG-1999; 99US-0151438;
PR 01-SEP-1999; 99US-0151930;
PR 07-SEP-1999; 99US-0152363;
PR 10-SEP-1999; 99US-0153070;
PR 13-SEP-1999; 99US-0153758;
PR 15-SEP-1999; 99US-0154018;
PR 16-SEP-1999; 99US-0154039;
PR 20-SEP-1999; 99US-0154779;
PR 22-SEP-1999; 99US-0155139;
PR 23-SEP-1999; 99US-0155486;
PR 24-SEP-1999; 99US-0155659;
PR 28-SEP-1999; 99US-0156458;
PR 29-SEP-1999; 99US-0156596;
PR 04-OCT-1999; 99US-0157117;
PR 05-OCT-1999; 99US-0157753;
PR 06-OCT-1999; 99US-0157865;
PR 07-OCT-1999; 99US-0158029;
PR 08-OCT-1999; 99US-0158232;
PR 12-OCT-1999; 99US-0158369;
PR 13-OCT-1999; 99US-0159293;
PR 13-OCT-1999; 99US-0159294;
PR 13-OCT-1999; 99US-0159295;
PR 14-OCT-1999; 99US-0159329;
PR 14-OCT-1999; 99US-0159330;
PR 14-OCT-1999; 99US-0159331;
PR 14-OCT-1999; 99US-0159637;
PR 14-OCT-1999; 99US-0159638;
PR 14-OCT-1999; 99US-0159584;
PR 18-OCT-1999; 99US-0160741;
PR 21-OCT-1999; 99US-0160767;
PR 21-OCT-1999; 99US-0160768;
PR 21-OCT-1999; 99US-0160770;
PR 21-OCT-1999; 99US-0160814;
PR 21-OCT-1999; 99US-0160815;
PR 22-OCT-1999; 99US-0160980;
PR 22-OCT-1999; 99US-0160981;
PR 22-OCT-1999; 99US-0160989;
PR 25-OCT-1999; 99US-0161404;
PR 25-OCT-1999; 99US-0161405;
PR 25-OCT-1999; 99US-0161406;
PR 26-OCT-1999; 99US-0161359;
PR 26-OCT-1999; 99US-0161360;
PR 26-OCT-1999; 99US-0161361;
PR 28-OCT-1999; 99US-0161920;
PR 28-OCT-1999; 99US-0161992;
PR 28-OCT-1999; 99US-0161993;
PR 29-OCT-1999; 99US-0162142;
Query Match 36.5%; Score 978.5; DB 21; Length 492;
Best Local Similarity 41.9%; Pred. No. 3.2e-76;
Matches 210; Conservative 79; Mismatches 171; Indels 41; Gaps 8;

Db 84 vaigsvfemthlfgeptlfsadpeneftvlqnegllfscspasicnllghkslllmgs 143
Qy 132 MHRMRSISLNFSLHARLRTLLKDVESHTEFLVDSWOONSIFSADPAKFTENLMAKH 191
Db 144 lhrmrslmsfmsnsllkchlmldldtvlrfnldswessrvll--meaakkftfelvkq 201
Qy 192 IMSMDGEEETOLKKEVTFMKGVVSAPLNLPGVAHAKQSRATILKFERKKEEKL 251
Db 202 lmsldpg-ewesekirkeyllvlegfslpflfstlytkalqar-----rkvaalt 252
Qy 252 DIKEEDQEEFEKTEDEEMSKSDHVRKQRTDDDLGWLKHNSLSTQILDJLISLFA 311
Db 253 vvmkkrreeeegea-----rkmdaalaaaddgsdeideivdvalilla 298
Qy 312 GHETSSVALALAIFFLQACPAVEELREHELEIARAKKELGESLMDYKMDFTQCVI 371
Db 299 gyetstlmtlavkfltetplalaqlkeehetiramed--syslewedyksmptqcv 356
Qy 372 NETLRIGWVVRFLHRKALKDVRKGYDIPSGMKVLPVISAHLDNSRYDQPLFPMRWQ 431
Db 357 netlrvanllgvtfrantdveikykikpgkwvfssfravldphnkdartltpwtwq 416
Qy 432 QONNGASSSGSFSFTWGN--YMPGCGPRLCAGSELAKLEMAVFIRHVLKFNELAE 490
Db 417 -----snsvltgspnvtftpggprlcpgeylarvalsvllhvlvgfswpaeg 466
Qy 491 DQFAFPFVDEPPNGLPIRVSR 511
Db 467 dkivffptctckryplfvkr 487

RESULT 7
AAG4572
ID AAG4572 standard; Protein: 444 AA.
AC AAG4572;
XX 18-OCT-2000 (first entry)
DT
XX Arabidopsis thaliana protein fragment SEQ ID NO: 55848.
DE
XX Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.
XX Arabidopsis thaliana.
OS
XX EP1033405-A2.
PN
XX
XX
PD 06-SEP-2000.
XX
PE 25-FEB-2000; 2000EP-0301439.
XX
PR 25-FEB-1999; 99US-0121825;
PR 05-MAR-1999; 99US-0123180;
PR 09-MAR-1999; 99US-0123548;
PR 23-MAR-1999; 99US-0125788;
PR 25-MAR-1999; 99US-0126264;
PR 29-MAR-1999; 99US-0126785;
PR 01-APR-1999; 99US-0127462;
PR 06-APR-1999; 99US-0128234;
PR 08-APR-1999; 99US-0128714;
PR 16-APR-1999; 99US-0129845;
PR 19-APR-1999; 99US-0130077;
PR 21-APR-1999; 99US-0130449;
PR 23-APR-1999; 99US-0130510;
PR 23-APR-1999; 99US-0130891;
PR 28-APR-1999; 99US-0131449;
PR 30-APR-1999; 99US-0132048;
PR 30-APR-1999; 99US-0132407;
PR 04-MAY-1999; 99US-0132484;
PR 05-MAY-1999; 99US-0132485;

PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138647.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139432.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139452.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.

PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156586.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.

RESULT 9
AAG46490
ID AAG46490 standard; Protein; 461 AA.
XX
AC AAG46490;
XX
DT 18-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana protein fragment SEQ ID NO: 58494.
XX
KW Protein identification; signal transduction pathway; metabolic pathway;
hybridisation assay; genetic mapping; gene expression control; promoter;
termination sequence.
XX
OS Arabidopsis thaliana.
XX
PN EPI033405-A2.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-0301439.
XX
PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 23-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 28-APR-1999; 99US-0130891.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136382.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139482.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.

PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 05-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.

```

PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161982.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

```

Query Match 27.1% Score 726; DB 21: Length 461;

Best Local Similarity 32.3%; Pred. NO. 6.6e-56; Matches 166; Conservative 91; Mismatches 181; Indels 76; Gaps 10;

```

QY 12 LLLPSLLSLFLILKRRN--RKTRENLPPKSGMPFLGETIGYKPYATATLGDQM 69
DB 7 LLLVLSLSEALL-----rvnqmytknglpypgcmgwpifgetefikqgp-----nfmr 55
QY 70 QHVSRYGKIYNSNLFGEPTIYSADAGLNRTLLQNGRLFECSYPSRIGCIIIGKWSMLVLV 129
DB 56 ngrlrygsffkshllgpcplismdsdevrlylkneskglvpypqsmldilgctmmaavh 115
QY 130 GDMHRDMSISLNFSLSHRLRTIILKQVERHTLFLVDSWQONSIFSODEKKKTF----- 185
DB 116 gsshrlnrgslslststcmmdhllpkvdfmrslqdwelvdldqdklhmaflssl 175

```

```

QY 186 -----NLMAKHMSMDPOEETEOUKKRYTFMKGVVASPLNPGTAVHKAQSRATLK 240
DB 176 tqiagnltpfv-----eeftafkflvgtlsvpldipgltncryrgidarmidr 225
QY 241 FIERKMEERKLIDKEEDOEDEEVKTEDEAEMSKSDHVRKQRTDDLGLWLKHSN---L 296
DB 226 llrelmqger-----dsge-----tfcdmlygmkkegncyp 258
QY 297 STEQILDLILSLFAGHTSSVAIALAIFLQACPKAVEELREHLEFLARKKELGESEL 356
DB 259 tdeairdqvvcllysgyeltvstsmmalkylhdpkqelraehlaefrekrq--depl 316
QY 357 NMDYKKKMDFTQCVINETRLRGVNVYRFLHRKALKDVRKGYDIFSGMKVLPVISAVHLDN 416
DB 317 gledvksmkftravilyetsrlativngylrtktrdleinnglllpkywrllyytreinyda 376
QY 417 SRYDQPLNFPMWQOONNGASSSGSFSFTWGNMYMPFGGPRLCAGSELAKLEMAVFI 476
DB 377 nlyedpllfnpwrwmkkslesq-----nscfvgggrtlcpgekigvelssfl 425
QY 477 HHVLKFMELAEDDQPRAPFVDFPNCPLIRVS 510
DB 426 hyfvtiryweigdelmvfrvflapkyfhlrls 459

RESULT 10
AAG46489
ID AAG46489 standard; Protein: 462 AA.
XX AC
XX AAG46489;
XX DT 18-OCT-2000 (first entry)
XX DE Arabidopsis thaliana protein fragment SEQ ID NO: 58493.
XX KW Protein identification; signal transduction pathway; metabolic pathway; hybridisation assay; genetic mapping; gene expression control; promoter; termination sequence.
XX OS Arabidopsis thaliana.
XX PN EP1033405-A2.
XX PD 06-SEP-2000.
XX PE 25-FEB-2000; 2000EP-0301439.
XX PR 25-FEB-1999; 99US-0121825.
XX PR 05-MAR-1999; 99US-0123180.
XX PR 09-MAR-1999; 99US-0123548.
XX PR 23-MAR-1999; 99US-0125788.
XX PR 25-MAR-1999; 99US-0126264.
XX PR 29-MAR-1999; 99US-0126785.
XX PR 01-APR-1999; 99US-0127462.
XX PR 06-APR-1999; 99US-0128234.
XX PR 08-APR-1999; 99US-0128714.
XX PR 16-APR-1999; 99US-0129845.
XX PR 19-APR-1999; 99US-0130077.
XX PR 21-APR-1999; 99US-0130449.
XX PR 23-APR-1999; 99US-0130510.
XX PR 28-APR-1999; 99US-0130891.
XX PR 30-APR-1999; 99US-0132048.
XX PR 30-APR-1999; 99US-0132407.
XX PR 04-MAY-1999; 99US-0132484.
XX PR 05-MAY-1999; 99US-0132485.
XX PR 06-MAY-1999; 99US-0132486.
XX PR 07-MAY-1999; 99US-0132487.
XX PR 11-MAY-1999; 99US-0132863.
XX PR 14-MAY-1999; 99US-0134218.
XX PR 14-MAY-1999; 99US-0134219.

```

PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.

PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 18-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.

PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161922.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 27.18; Score 726; DB 21; Length 462;
Best Local Similarity 32.38; Pred. No. 6,6e-56;

Matches 166; Conservative 91; Mismatches 181; Indels 76; Gaps 10;

QY 12 LLLPSLLSLFLFLRLRN--RKTRFNLPGKSGMPELSEETGYLKYATATLGDMDQ 69
DB 8 LLLVLSLALL-----rwmgytknglppgtmgwplfgetlefkpgp-----ntmr 56
QY 70 QHVSRYKRIYSNLFEGEPTIVSADAGLNRFILQNEGRLECSYPRISIGILGKSMELV 129
DB 57 ngrlrygffkshllgpcflismdeevnyllkneskglvpqysmdilgtcmaavh 116
QY 130 GDMHRDMSISLNLFLSHARKLRTILKDYERHTLPVLDWMQNSITFSADDEAKRTF--- 185
DB 117 gssnrlmgsllstlssctmmrdhllpkvdhmrsvldqwnelavldqdkckhmaflssl 176
QY 186 -----NLMAKHMSMDPCEETEQLEKREYTEMKGVASAPLNLPQTAVYHKAQSRATILK 240
DB 177 tqiagnllkprty-----eeftakfklvgtlsvpldipgtnyrgsqidatnmidr 226
QY 241 FIERKMEERKLDIEEDQEEVEKTEDEAKNSKSDHVKORTDDLLGWLKHSN---L 296
DB 227 llreimgerr-----dege-----tfcmlygylmkkegnrypl 259
QY 297 STEQILDILSLFRAGHTSSVATAALAFPLQACPKAYEUREHELETLARKKKEGSEL 356
DB 260 ldeetrdgvlllysgyeltstsmaklylndhpkalqelraefrekrq--depl 317
QY 357 NMDDYKKKMDFTQCIYNETLRAGNVVRFLLRKALKDVRKYGYDIPSGWVLPVISAVIDN 416
DB 318 gledkxsmkfravlyetsrlatlvngylrkttrdlnngyllpkgyviyyteinyda 377
QY 417 SRYDQNLFNPMWMOQONNGASSSGSGSFTWGNVMPFGGPGPRLCAGSELAKLEMAVEI 476
DB 378 nlycedpplfnpmwrmkksleesg-----nscfvgggtlrcpckelgelveisfl 426
QY 477 HHLVKEWMLAEDDQPAFPFVDFPNGLPTRVS 510
DB 427 hyflvryrweelggdelmvrprvfapkgfhlrls 460

RESULT 11

AAAG46491
ID AAAG46491 standard; Protein: 465 AA.

AC AAAG46491;

DT 18-OCT-2000 (first entry)

DE Arabidopsis thaliana protein fragment SEQ ID NO: 58495.

KW Protein identification; signal transduction pathway; metabolic pathway;

KW hybridisation assay; genetic mapping; gene expression control; promoter;

KW termination sequence.

OS Arabidopsis thaliana.

PN EPI033405-A2.

PD 06-SEP-2000.

PF 25-FEB-2000; 2000EP-0301439.

PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 28-APR-1999; 99US-0130891.
PR 30-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 04-MAY-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 07-MAY-1999; 99US-0132487.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137529.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142380.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.

Db 430 hYfVtYrweei9gdElmVfprVfapKgfhlrls 463
RESULT 12
AA011836 standard; Protein: 461 AA.
XX AA011836;
XX 17-OCT-2000 (first entry)
XX Arabidopsis thaliana protein fragment SEQ ID NO: 10708.
XX Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
XX termination sequence.
XX Arabidopsis thaliana.
XX EP1033405-A2.
XX
XX 06-SEP-2000.
XX
XX 25-FEB-2000; 2000EP-0301439.
XX
XX 25-FEB-1999; 99US-0121825.
XX 05-MAR-1999; 99US-0123180.
XX 09-MAR-1999; 99US-0123548.
XX 23-MAR-1999; 99US-0125788.
XX 23-MAR-1999; 99US-0126284.
XX 29-MAR-1999; 99US-0126785.
XX 01-APR-1999; 99US-0127462.
XX 06-APR-1999; 99US-0128234.
XX 08-APR-1999; 99US-0128714.
XX 16-APR-1999; 99US-0129845.
XX 19-APR-1999; 99US-0130079.
XX 21-APR-1999; 99US-0130449.
XX 23-APR-1999; 99US-0130510.
XX 23-APR-1999; 99US-0130891.
XX 28-APR-1999; 99US-0131449.
XX 30-APR-1999; 99US-0132048.
XX 30-APR-1999; 99US-0132407.
XX 04-MAY-1999; 99US-0132484.
XX 05-MAY-1999; 99US-0132485.
XX 06-MAY-1999; 99US-0132486.
XX 06-MAY-1999; 99US-0132487.
XX 07-MAY-1999; 99US-0132863.
XX 11-MAY-1999; 99US-0134256.
XX 14-MAY-1999; 99US-0134218.
XX 14-MAY-1999; 99US-0134219.
XX 14-MAY-1999; 99US-0134221.
XX 14-MAY-1999; 99US-0134370.
XX 18-MAY-1999; 99US-0134768.
XX 19-MAY-1999; 99US-0134941.
XX 20-MAY-1999; 99US-0135124.
XX 21-MAY-1999; 99US-0135353.
XX 24-MAY-1999; 99US-0135629.
XX 25-MAY-1999; 99US-0136021.
XX 27-MAY-1999; 99US-0136382.
XX 28-MAY-1999; 99US-0136782.
XX 01-JUN-1999; 99US-0137222.
XX 03-JUN-1999; 99US-0137528.
XX 04-JUN-1999; 99US-0137502.
XX 07-JUN-1999; 99US-0137724.
XX 08-JUN-1999; 99US-0138094.
XX 10-JUN-1999; 99US-0138540.
XX 10-JUN-1999; 99US-0138847.
XX 14-JUN-1999; 99US-0139119.
XX 16-JUN-1999; 99US-0139452.
XX 16-JUN-1999; 99US-0139453.
XX 17-JUN-1999; 99US-0139492.
XX 18-JUN-1999; 99US-0139454.
XX 18-JUN-1999; 99US-0139455.

PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140981.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142380.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144335.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144684.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145102.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 26-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 28-JUL-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.

```
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160747.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161983.
PR 29-OCT-1999; 99US-0162142.

Query Match 26.6%; Score 712; DB 21; Length 461;
Best Local Similarity 32.5%; Pred. No. 1,1e-54;
Matches 164; Conservative 93; Mismatches 196; Indels 52; Gaps 8;
```

```
Db 116 gshrlmrgslslslspsrmkdhllpkiddfmrnlylgwddletvdqekthmaif-1ss 174
Qy 190 KHIMSDPGEETEOJLKEEYTFMKGVASAPLNPGYAHKALOSRAPIILKEIKMEER 249
Db 175 llqiaetlkpeveeyreteffklvgtlsvpidpgtnyrsvgarmidrlitelmqer 234
Qy 250 KLIDKEEDQEEFEVYTEDEAEKMSDHYRKQRTDDDLGWIYKHNN----LSTEDILDI 305
Db 235 k-----esgetft-----dmlylkkednrylltdkeirgqv 267
Qy 306 LSILFAGHETSSVALAIAIFLOACPKAVEELREEHLELAKAKKGESELMWDYKMD 365
Db 268 vtllysgyetvstcmmaikylhdpkpkaleelrrehlairekrp--depltdidkismk 325
Qy 366 FTQCVINETLRGNVRFPLHKKALKDVRKGYDIPSGMKVLPVISAHLNDSRYDQPIIF 425
Db 326 flravifetsrlativngvlrktthdlnyglipkgywilyytrelnydselyedpmif 385
Qy 426 NPWRMOQONNGASSSGSFSFTWGNMYMPFGGPRLAGSELAKEIMAVFIHHLKRNW 485
Db 386 npwrmekslssks-----yflfggvrllcpqkelgisevsflnyftvktyw 434
Qy 486 ELAEDDQFAFPFVDFPGLPIRVS 510
Db 435 geengedklmfvrsapkyhikcs 459
```

```
RESULT 13
AAG11835
ID AAG11835 standard; Protein; 462 AA.
XX
AC AAG11835;
XX
DT 17-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana protein fragment seq ID NO: 10707.
KW Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
XX termination sequence.
XX OS Arabidopsis thaliana.
XX PN EPI033405-A2.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-0301439.
XX
PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 04-MAY-1999; 99US-0132407.
PR 05-MAY-1999; 99US-0132484.
PR 06-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 07-MAY-1999; 99US-0132487.
PR 11-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
```

PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134376.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140981.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.

PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 03-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159684.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.

PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 26-OCT-1999; 99US-0161362.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 26.6%; Score 712; DB 21; Length 462;
Best Local Similarity 32.5%; Pred. No. 1,1e-54;

Matches 164; Conservative 93; Mismatches 196; Indels 52; Gaps 8;

QY 12 LLLPSLLSLFLITLKRRL-RKTRPNLPKSGWFLGCTGYLKPYATLTGDMQ 69
DB 2 mmllglvllvcltallrwgmtysskqylppjgmwplfgeteflkqgp----dfmk 56
QY 70 QHVKYKGIKYSNLFGEPTIYSADAGINRFLIOWNGRLFECSYPSIGCTIGKMSMLV 129
DB 57 nqrlrsgfifshllgpcrlvsmcaelrlylmeskylvagypqsmldllgtcliaavh 116
QY 130 GDMHRMRSISLNFSLHARLTILKDERHTLFLVDSMOONSIFSADKAKKFTFNMA 189
DB 117 gpsnrlmgslslslspmmkdhllpkldfmrnylsgwddletvldqektkmaf-iss 175
QY 190 KHISMDBGEETQOLKEVYTFMKGVASAPLNLPGTAHKALOSRAITLKFIERKMEER 249
DB 176 llqlaeltkpveeyrtefkflvgtlslvpldipgtnlysgvqnmldrltelmger 235
QY 250 KLDIKEDEQEEFEKTEDEAMSKSDHAKQRTDDLLGNLKHNSN---LSTEDILDI 305
DB 236 k-----esgeft-----dmlygmkkednrylltdkeirgv 268
QY 306 LSLFAGHETSSVAIALAIFLQACPKAVEELREHELEIAAKKELGESELMWDYKRM 365
DB 269 vtllysgyetsvtcsmalkylhdpkaleelrcheiairckrp--depltdidkismk 326
QY 366 FTQCVINETLGNVFLHRAKALKDVRKGYDIPSGMKVLPVISAHLDSRYDQPMLE 425
DB 327 ftravifetsrlatlngvllktchdelngyllpkgyrlyvvtrelngydtlsyedmllf 386
QY 426 NPMWMOQNNNGSSGSFSTGWNVMPRGGPRLCAGSELAKEMAVFIHHLVLEKFM 485
DB 387 npwtmekslesks-----yillfggyrlcpqkelglssevsfihyftkyrw 435
QY 486 ELAEDDOPFAFPYDFPNGLPIRVS 510
DB 436 eengedklmvfprvsapkyhks 460

RESULT 14
AAG1834
ID AAG1834 standard; Protein: 465 AA.

AC AAG1834;

DT 17-OCT-2000 (first entry);

DE Arabidopsis thaliana protein fragment SEQ ID NO: 10706.

KW Protein identification: signal transduction pathway; metabolic pathway;
hybridisation assay; genetic mapping; gene expression control; promoter;
termination sequence.

OS Arabidopsis thaliana.

PN EP1033405-A2.

PD 06-SEP-2000.

XX

PF 25-FEB-2000; 2000EP-0301439.
XX 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123160.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134766.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139465.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139889.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140699.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.

OY 484 NWELAEEDDPFAFFVDFPNGLPPIRVS 510
Db 437 rweengedkilmvfrvsapsyhlkcs 463

RESULT 15
AAG20783

ID AAG20783 standard; Protein: 468 AA.

XX AAG20783;

XX 17-OCT-2000 (first entry)

XX Arabidopsis thaliana protein fragment SEQ ID NO: 23102.

KM Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
termination sequence.

XX Arabidopsis thaliana.

XX EP1033405-A2.

XX PD 06-SEP-2000.

XX 25-FEB-2000; 2000EP-0301439.

XX 25-FEB-1999; 99US-0121825.

PR 05-MAR-1999; 99US-0123180.

PR 09-MAR-1999; 99US-0123548.

PR 23-MAR-1999; 99US-0125788.

PR 25-MAR-1999; 99US-0126284.

PR 29-MAR-1999; 99US-0126785.

PR 01-APR-1999; 99US-0127462.

PR 06-APR-1999; 99US-0128234.

PR 16-APR-1999; 99US-0128714.

PR 19-APR-1999; 99US-0129845.

PR 21-APR-1999; 99US-0130077.

PR 23-APR-1999; 99US-0130449.

PR 23-APR-1999; 99US-0130510.

PR 28-APR-1999; 99US-0130891.

PR 30-APR-1999; 99US-0131449.

PR 30-APR-1999; 99US-0132048.

PR 04-MAY-1999; 99US-0132407.

PR 05-MAY-1999; 99US-0132484.

PR 06-MAY-1999; 99US-0132485.

PR 06-MAY-1999; 99US-0132486.

PR 07-MAY-1999; 99US-0132487.

PR 11-MAY-1999; 99US-0132863.

PR 14-MAY-1999; 99US-0134218.

PR 14-MAY-1999; 99US-0134219.

PR 14-MAY-1999; 99US-0134221.

PR 18-MAY-1999; 99US-0134370.

PR 18-MAY-1999; 99US-0134768.

PR 19-MAY-1999; 99US-0134941.

PR 20-MAY-1999; 99US-0135124.

PR 21-MAY-1999; 99US-0135353.

PR 24-MAY-1999; 99US-0135629.

PR 25-MAY-1999; 99US-0136021.

PR 27-MAY-1999; 99US-0136392.

PR 28-MAY-1999; 99US-0136782.

PR 01-JUN-1999; 99US-0137222.

PR 03-JUN-1999; 99US-0137528.

PR 04-JUN-1999; 99US-0137502.

PR 07-JUN-1999; 99US-0137724.

PR 08-JUN-1999; 99US-0138094.

PR 10-JUN-1999; 99US-0138540.

PR 10-JUN-1999; 99US-0138847.

PR 14-JUN-1999; 99US-0139119.

PR 16-JUN-1999; 99US-0139452.

PR 16-JUN-1999; 99US-0139453.

PR 17-JUN-1999; 99US-0139492.

PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147132.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 17, 2001, 19:19:21 : Search time 26.13 Seconds
(without alignments)
404.242 Million cell updates/sec

Title: US-09-502-426-2

Perfect score: 2681
Sequence: 1 MEETHEHLLPLLLPILLS.....FAFPVDPNGLPINVSRL 513

Scoring table: BLOSUM62
Gapop 10.0, Gapext 0.5

Searched: 197339 seqs, 20590346 residues

Total number of hits satisfying chosen parameters: 197339

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database:

Issued_Patents_AA:*
1: /cgn2_6/prodata/2/1aa/5A_COMB.pep:*
2: /cgn2_6/prodata/2/1aa/5B_COMB.pep:*
3: /cgn2_6/prodata/2/1aa/6A_COMB.pep:*
4: /cgn2_6/prodata/2/1aa/6B_COMB.pep:*
5: /cgn2_6/prodata/2/1aa/PCPUS_COMB.pep:*
6: /cgn2_6/prodata/2/1aa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description |
|------------|-------|-------------|--------|----|-------------------|
| 1 | 978.5 | 36.5 | 472 | 2 | US-08-622-166A-2 |
| 2 | 978.5 | 36.5 | 472 | 2 | US-08-622-166A-4 |
| 3 | 503 | 18.8 | 492 | 3 | US-08-724-466B-2 |
| 4 | 439 | 16.4 | 497 | 3 | US-08-724-466B-4 |
| 5 | 289 | 10.8 | 511 | 4 | US-08-991-677-4 |
| 6 | 279.5 | 10.4 | 513 | 3 | US-08-948-564-6 |
| 7 | 264 | 9.8 | 520 | 2 | US-09-091-432-2 |
| 8 | 263.5 | 9.8 | 504 | 1 | US-08-457-274A-25 |
| 9 | 263.5 | 9.8 | 504 | 5 | PCT-US95-05758-25 |
| 10 | 257 | 9.6 | 476 | 1 | US-08-313-075A-30 |
| 11 | 256.5 | 9.6 | 480 | 1 | US-08-201-118-7 |
| 12 | 256.5 | 9.6 | 480 | 2 | US-08-238-821B-7 |
| 13 | 256.5 | 9.6 | 480 | 5 | PCT-US95-05744-7 |
| 14 | 245 | 9.1 | 495 | 1 | US-08-532-065B-2 |
| 15 | 244.5 | 9.1 | 576 | 3 | US-08-948-564-16 |
| 16 | 241 | 9.0 | 510 | 3 | US-08-606-505B-66 |
| 17 | 241 | 9.0 | 510 | 4 | US-09-616-890-66 |
| 18 | 240.5 | 9.0 | 509 | 3 | US-08-948-564-18 |
| 19 | 239 | 8.9 | 496 | 1 | US-08-313-075A-50 |
| 20 | 236 | 8.8 | 521 | 3 | US-08-948-564-14 |
| 21 | 234.5 | 8.7 | 490 | 1 | US-08-201-118-1 |
| 22 | 234.5 | 8.7 | 480 | 2 | US-08-238-821B-1 |
| 23 | 234.5 | 8.7 | 480 | 5 | PCT-US95-05744-1 |
| 24 | 233.5 | 8.7 | 508 | 4 | US-08-991-677-2 |
| 25 | 231 | 8.6 | 506 | 1 | US-08-313-075A-38 |
| 26 | 231 | 8.6 | 506 | 3 | US-08-606-505B-65 |
| 27 | 231 | 8.6 | 506 | 4 | US-09-616-990-65 |

| | | | | | | |
|----|-------|-----|-----|---|-------------------|-------------------|
| 28 | 223.5 | 8.3 | 496 | 4 | US-09-292-768-64 | Sequence 64, Appl |
| 29 | 223 | 8.3 | 490 | 1 | US-08-201-118-3 | Sequence 3, Appl |
| 30 | 223 | 8.3 | 490 | 2 | US-08-238-821B-3 | Sequence 3, Appl |
| 31 | 223 | 8.3 | 490 | 5 | PCT-US95-05744-3 | Sequence 3, Appl |
| 32 | 222 | 8.3 | 490 | 1 | US-08-201-118-9 | Sequence 9, Appl |
| 33 | 222 | 8.3 | 490 | 2 | US-08-238-821B-9 | Sequence 9, Appl |
| 34 | 222 | 8.3 | 490 | 5 | PCT-US95-05744-9 | Sequence 9, Appl |
| 35 | 222 | 8.3 | 516 | 3 | US-08-948-564-12 | Sequence 12, Appl |
| 36 | 221.5 | 8.3 | 500 | 2 | US-08-314-601-2 | Sequence 2, Appl |
| 37 | 221.5 | 8.3 | 500 | 5 | PCT-US95-13051-2 | Sequence 2, Appl |
| 38 | 218 | 8.1 | 510 | 3 | US-08-948-564-4 | Sequence 4, Appl |
| 39 | 215 | 8.0 | 500 | 3 | US-08-881-784-9 | Sequence 9, Appl |
| 40 | 214.5 | 8.0 | 496 | 3 | US-08-881-784-1 | Sequence 1, Appl |
| 41 | 214.5 | 8.0 | 496 | 4 | US-09-292-768-2 | Sequence 2, Appl |
| 42 | 214.5 | 8.0 | 496 | 4 | US-09-292-768-66 | Sequence 66, Appl |
| 43 | 214 | 8.0 | 490 | 1 | US-08-201-118-13 | Sequence 13, Appl |
| 44 | 214 | 8.0 | 490 | 2 | US-08-238-821B-13 | Sequence 13, Appl |
| 45 | 214 | 8.0 | 490 | 5 | PCT-US95-05744-13 | Sequence 13, Appl |

ALIGNMENTS

RESULT 1
US-08-622-166A-2
Sequence 2, Application US/08622166A
Patent No. 5952545
GENERAL INFORMATION:
APPLICANT: KONCZ, CSABA
APPLICANT: MATHUR, JAIDEEP
APPLICANT: SZEKERES, MIKLOS
APPLICANT: ALTMANN, THOMAS
TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING CYTOCHROME
TITLE OF INVENTION: P450-TYPE PROTEINS INVOLVED IN THE BRASSINOSTEROID SYNTHESIS
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESSEE: Birch, Stewart, Kolasch & Birch, LLP
STREET: P.O. Box 747
CITY: Falls Church
STATE: Virginia
COUNTRY: USA
ZIP: 22040-0747
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30 (EPO)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/622,166A
FILING DATE: 27-MAR-1996
CLASSIFICATION: 800
ATTORNEY/AGENT INFORMATION:
NAME: SVENSSON, LEONARD R.
REGISTRATION NUMBER: 30,330
REFERENCE/DOCKET NUMBER: 0147-0153P
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 205-8000
TELEFAX: (703) 205-8050
TELEX: 248345
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 472 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-622-166A-2

Query Match 36.5%; Score 978.5; DB 2; Length 472;
Best Local Similarity 41.9%; Pred. No. 2.1e-89;
Matches 210; Conservative 79; Mismatches 171; Indels 41; Gaps 8;
QY 12 LLLPSSLILLPLLLKRRNRKTRFNLPKSGMPFLGRTIGYLRKPYTATTLDPMQOH 71


```

: APPLICATION NUMBER: 08/667,546
: FILING DATE: June 21, 1996
: ATTORNEY/AGENT INFORMATION:
: NAME: Hunt, John C.
: REGISTRATION NUMBER: 36,424
: REFERENCE/DOCKET NUMBER: 50767/00004
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (416) 863-4344
: TELEFAX: (416) 863-2653
: INFORMATION FOR SEQ ID NO: 2:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 492 amino acids
: TYPE: amino acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: US-08-724-466B-2

Query Match      18.8%; Score 503; DB 3; Length 492;
Best Local Similarity 27.6%; Pred. No. 8,4e-42;
Matches 144; Conservative 95; Mismatches 216; Indels 66; Gaps 14;

Oy 7 HTLLPLLLPSLSLLFL-----ILKRRNRKTRFNLPPKSGMPFLGETIGYLPK 58
   ||| : : : ||| : : : ||| : : : ||| : : : ||| : : : ||| : : :
Db 4 YTLAVTFELCTIVLPLVLLFLAIVKLMELMIRVDPCRSPLPPTGMDLPFGETIGLI-- 61
   ||| : : : ||| : : : ||| : : : ||| : : : ||| : : : ||| : : :

Oy 59 YTATTLDPMOQHSKYGYKIRSNLFGPTIVSADAGLNRFILONEGRLEFCSYPSISIG 118
   ||| : : : ||| : : : ||| : : : ||| : : : ||| : : : ||| : : :
Db 62 ---LQRRKFLMKRKQKIGCYIKTHLFGNPTVRVGADNVRKILGHEHLVSVMWPAVSRT 118
   ||| : : : ||| : : : ||| : : : ||| : : : ||| : : : ||| : : :

Oy 119 ILGKWSMLVYLVGDHNRD-----MRSISLNFSLHARLRTILKDYERHRLFVLDGW- 172
   ||| : : : ||| : : : ||| : : : ||| : : : ||| : : : ||| : : :
Db 119 ILGSDTLNSVNGVQHKKKKAIMAFSDALEH-----YIPVQGEKSAIQEMLOKDS 172
   ||| : : : ||| : : : ||| : : : ||| : : : ||| : : : ||| : : :

Oy 173 IFSADDEAKKTFNLMAMHNSMDPGEET--EOLKEEYVTFMGVSAAPLNPCTAYHK 230
   ||| : : : ||| : : : ||| : : : ||| : : : ||| : : : ||| : : :
Db 173 CVLVYPEKKKLMFRIAMRILLEGPEPOIKTDEOLEVEAFEEIMIKLFLSPIDVPSGLYR 232
   ||| : : : ||| : : : ||| : : : ||| : : : ||| : : : ||| : : :

Oy 231 ALQSRATILKFERKMERKLDIKEEDQEEEVKTEDEAEKSKSHVAKQRTDDDLGW 290
   ||| : : : ||| : : : ||| : : : ||| : : : ||| : : : ||| : : :
Db 233 GLRRK---NFIHSKIEENIRKKIIDDNDNEQYKDALQLL---LNSRSRSD----- 279
   ||| : : : ||| : : : ||| : : : ||| : : : ||| : : : ||| : : :

Oy 291 LKHSNLTSTEOILDLILSLFAGHETSSVAIALAIFLQACPKAVEELREEHLEIRAKE 350
   ||| : : : ||| : : : ||| : : : ||| : : : ||| : : : ||| : : :
Db 280 ---PFSIQAKKEATETELFGGHETTASTSLVWFLGLNVEVQKVEB---VQEKVE 331
   ||| : : : ||| : : : ||| : : : ||| : : : ||| : : : ||| : : :

Oy 351 LG---ESELNMDYKKMDFTQCVINETLRLGNVVRFLHRAKALDVRYKKGYDIPSGMYL 406
   ||| : : : ||| : : : ||| : : : ||| : : : ||| : : : ||| : : :
Db 332 MGNVTPGKGLSMELLDOLKYGCVIKETLRINPVPVGGFRVALKTFELNGYQIPIKGMVVI 391
   ||| : : : ||| : : : ||| : : : ||| : : : ||| : : : ||| : : :

Oy 407 PVISAVHLDNSRYQOPNLFNPMWMOQONNGASSSGSFSFTWGNMYMPFGGPRICASE 466
   ||| : : : ||| : : : ||| : : : ||| : : : ||| : : : ||| : : :
Db 392 YSIDTHVADVFPNKEEFQPERF-----SKLEDGSRF--NYIFPGGGRMCVKE 442
   ||| : : : ||| : : : ||| : : : ||| : : : ||| : : : ||| : : :

Oy 467 LAKLEMAVFIHHLVLFKFMELAEDDQ-----PFAFPVDFP 502
   ||| : : : ||| : : : ||| : : : ||| : : : ||| : : : ||| : : :
Db 443 FAKVLKIFLVELLQHCNWLISNGPPTKGTPTIYPVUNLP 483
   ||| : : : ||| : : : ||| : : : ||| : : : ||| : : : ||| : : :

RESULT 4
US-08-724-466B-4
: Sequence 4, Application US/08724466B
: Patent No. 6063606
: GENERAL INFORMATION:
: APPLICANT: Petkovich, P. Martin, White, Jay A.,
: APPLICANT: Beckovitch, Barbara R., Jones, Glenaville
: TITLE OF INVENTION: Retinoid Metabolizing Protein
: NUMBER OF SEQUENCES: 30
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Blake, Cassels & Graydon
: STREET: Box 25, Commerce Court West
: CITY: Toronto
: ZIP: M5L 1A9

```

```

: COUNTRY: Canada
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Diskette, 3 1/2 inch, 1.4 Mb storage
: COMPUTER: COMPAQ, IBM PC compatible
: OPERATING SYSTEM: MS-DOS 5.1
: SOFTWARE: WORD PERFECT
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/724,466B
: FILING DATE: October 1, 1996
: PRIORITY APPLICATION DATA:
: APPLICATION NUMBER: 08/667,546
: FILING DATE: June 21, 1996
: ATTORNEY/AGENT INFORMATION:
: NAME: Hunt, John C.
: REGISTRATION NUMBER: 36,424
: REFERENCE/DOCKET NUMBER: 50767/00004
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (416) 863-4344
: TELEFAX: (416) 863-2653
: INFORMATION FOR SEQ ID NO: 4:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 497 amino acids
: TYPE: amino acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: US-08-724-466B-4

Query Match      16.4%; Score 439; DB 3; Length 497;
Best Local Similarity 27.0%; Pred. No. 2,2e-35;
Matches 142; Conservative 85; Mismatches 226; Indels 72; Gaps 15;

Oy 10 LPULLPLBL-----LSLLFLILK-----RRNRKTRFNLPPKSGMPFLGETIGYLPK 57
   ||| : : : ||| : : : ||| : : : ||| : : : ||| : : : ||| : : :
Db 3 LPALLASLCTFVPLPLFLFLAIIKIMLDLYCVSGRDSALPLPPTMGEPFPEFTL--- 58
   ||| : : : ||| : : : ||| : : : ||| : : : ||| : : : ||| : : :

Oy 58 PYTATTLDPMOQHSKYGYKIRSNLFGPTIVSADAGLNRFILONEGRLEFCSYPSISIG 117
   ||| : : : ||| : : : ||| : : : ||| : : : ||| : : : ||| : : :
Db 59 -QMVLRKRFLOMKRRKRGFYIKTHLFGRPYVRVGADNVRKILGDDRLVSVMWPAVSIR 117
   ||| : : : ||| : : : ||| : : : ||| : : : ||| : : : ||| : : :

Oy 118 GLGKWSMLVYLVGDHNRDMSISLNFSLHARL--TILKDYERHRLFVLDGWQO--NS 172
   ||| : : : ||| : : : ||| : : : ||| : : : ||| : : : ||| : : :
Db 118 TLISGSCSLNLDSSHOKRKVIMRAFSEALECYVPVITEVGSS---LEQWLSGCR 173
   ||| : : : ||| : : : ||| : : : ||| : : : ||| : : : ||| : : :

Oy 173 IFSADDEAKKTFNLMAMHNSMDP-----GRETEOLKKEIVTFMGVSAAPLNPCTA 227
   ||| : : : ||| : : : ||| : : : ||| : : : ||| : : : ||| : : :
Db 174 GLVYPEVKRILMFRIAMRILLEGCEPOLAGDGDSE--QOLVEAFEEIMTRNLFSLPIDVPSG 232
   ||| : : : ||| : : : ||| : : : ||| : : : ||| : : : ||| : : :

Oy 228 YKALQSRATILKFERKMERKLDIKEED-----QEEEVKTEDEAEKSKSDHYRKORT 282
   ||| : : : ||| : : : ||| : : : ||| : : : ||| : : : ||| : : :
Db 223 LYRGKAKARLHARLEONIRAKTICGLRASGCGCKDALQILIEHSW-----RGERL 285
   ||| : : : ||| : : : ||| : : : ||| : : : ||| : : : ||| : : :

Oy 283 DDDLGLWVLLKHSNLTSTEOILDLILSLFAGHETSSVAIALAIFLQACPKAVEELREHL 342
   ||| : : : ||| : : : ||| : : : ||| : : : ||| : : : ||| : : :
Db 286 DMQ-----ALKQS--STE-----LFGGHETTASATSLITVYGLVPHVLOKVBELK 331
   ||| : : : ||| : : : ||| : : : ||| : : : ||| : : : ||| : : :

Oy 343 ELARAKKEGESELMWDDYKKMDFTQCVINETLRLGNVVRFLHRAKALDVRYKKGYDIPSG 402
   ||| : : : ||| : : : ||| : : : ||| : : : ||| : : : ||| : : :
Db 332 SKGLCKSNODKRLMEILLOLYIGCVIKETLRINPVPVGGFRVALKTFELNGYQIPIK 391
   ||| : : : ||| : : : ||| : : : ||| : : : ||| : : : ||| : : :

Oy 403 WKVLPVISAVHLDNSRYQOPNLFNPMWMOQONNGASSSGSFSFTWGNMYMPFGGPRIC 462
   ||| : : : ||| : : : ||| : : : ||| : : : ||| : : : ||| : : :
Db 392 MNVITSIDTHVAVIEFNKKEEFNDRSAPHPEDASRPS-----FTFPGGGLKSC 442
   ||| : : : ||| : : : ||| : : : ||| : : : ||| : : : ||| : : :

Oy 463 AGSELAKLEMAVFIHHLVLFKFMELAE-----DDQPFAPFVDFP 502
   ||| : : : ||| : : : ||| : : : ||| : : : ||| : : : ||| : : :
Db 443 VKEFAKILKIFLVELARHCDWQLLNGPPTKGTPTIYPVUNLP 487
   ||| : : : ||| : : : ||| : : : ||| : : : ||| : : : ||| : : :

RESULT 5
US-08-991-677-4
: Sequence 4, Application US/08991677A
: Patent No. 6252135

```

```

: GENERAL INFORMATION:
: APPLICANT: Chiang, Vincent L
: APPLICANT: Carraway, Daniel T
: APPLICANT: Smeltzer, Richard H
: TITLE OF INVENTION: Production of Syngly lignin in Gymnosperms
: FILE REFERENCE: 50617
: CURRENT APPLICATION NUMBER: US/08/991.677A
: CURRENT FILING DATE: 1997-12-16
: EARLIER APPLICATION NUMBER: US 60/033,381
: EARLIER FILING DATE: 1996-12-16
: NUMBER OF SEQ ID NOS: 11
: SOFTWARE: Patent In Ver. 2.0
: SEQ ID NO 4
: LENGTH: 511
: TYPE: PRT
: ORGANISM: Liquidambar styraciflua
: US-08-991-677-4

```

```

Query Match      10.8%; Score 289; DB 4; Length 511;
Best Local Similarity 22.0%; Pred. No. 2,3e-20;
Matches 121; Conservative 95; Mismatches 196; Indels 138; Gaps 21;

```

```

QY 6 HHTLLPL-LLLPSLLSLFLILKRRNRKTRNLPKSGMPFLGTTIGYLKPYTATTL 64
D 6 HEALQPLPMLTFLIPLILGLVSR--LRQLPYPPKGLPIVGNML--MMDLTHHGL 62
QY 65 GDFMOHVSXKGIYRSNLFGEPTIVSADAGLNREIIONERLFE-----CSFPR 114
D 63 AKLAKQ-----YGLFLHKMGFLHMAVSTPMAROVLOVONIFSNRPATTAISLYYDR 118
QY 115 SIGGILGKWSMLVGVDMHRMRSIS--LNFLSHARLFTILKDVERTHLEFLVDSMOQMSI 173
D 119 AD-----MAFAHYGPPWRQMRKLCVMKLESRRK-----AESME----- 151
QY 174 FSAODE-----AKFTFNLMKHHMSMDPGF-----EETE 203
D 152 -SYRDEVDASAVVAVASNGSTVNIQELVFLTKNITVRAFGTISHEDODEFVAILOEFS 210
QY 204 QL-----KREVTETMKGV--VSAPLNLPRTAYHKALOSRAITILFERKKEERKLDIKE 255
D 211 QLFQAFNINADIPMLKVPVPGINVRN-----KARALDGFIDKIIDH---IQK 257
QY 256 EDQEEVEKTE-----DEAEMSKSDHVRKQRTDDLLGWLKHSNLSTEQIIDLI 305
D 258 GSKNSEEVDTMDVDDLAFYGEAKVSESD-----LQNSILTLTNDNIKAI 303
QY 306 LSLFAGHETSSVAIALAIFLQACPKAVELEBREHLEIARAKKELGSEELNMDYKMD 365
D 304 MDVMEGTEIVASAIEMAMELKMSPEDLKVKQOELAVVVGIDRRVEEK-----DFEKL 358
QY 366 FTQCVINETLGLVNVREFLHRKALKDVRKGYDIPSGMKVLPVISAHLDSRYDOPNLF 425
D 359 YLKCVLEVLRLHPILLHETRAEDAEVGGYITPAKSRVINACALGRDNSADPDTF 418
QY 426 NPWMOQONNGASSGSGSFSTWGN--YMPFGGPRLCASELAKEMAVFIHHLVLKF 483
D 419 RPSPEFLK-----DGVPDFK--GNNEFEIIFGSGRSCSPGMQGLAVLETTVAHLHCF 469
QY 484 NWELEADDP 493
D 470 TWELPDGMRP 479

```

```

RESULT 6
US-08-948-564-6
: Sequence 6, Application US/08948564
: Patent No. 6121512
: GENERAL INFORMATION:
: APPLICANT: Siminszky, Balazs
: APPLICANT: Dewey, Ralph E.
: APPLICANT: Cordin, Frederick T.
: TITLE OF INVENTION: No. 6121512e1 Cytochrome P-450 Constructs and

```

```

: TITLE OF INVENTION: Methods of Producing Herbicide-Resistant Transgenic Plants
: NUMBER OF SEQUENCES: 23
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Virginia C. Bennett
: STREET: PO Box 37428
: CITY: Raleigh
: STATE: No. 6121512th Carolina
: COUNTRY: USA
: ZIP: 27627
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patent In Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/948,564
: FILING DATE:
: CLASSIFICATION: 800
: ATTORNEY/AGENT INFORMATION:
: NAME: Bennett, Virginia C.
: REGISTRATION NUMBER: 37,092
: REFERENCE/DOCKET NUMBER: 5051-409
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 919-854-1401
: TELEFAX: 919-854-1401
: INFORMATION FOR SEQ ID NO: 6:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 513 amino acids
: TYPE: amino acid
: TOPOLOGY: linear
: MOLECULE TYPE: protein
: US-08-948-564-6

```

```

Query Match      10.48; Score 279.5; DB 3; Length 513;
Best Local Similarity 22.9%; Pred. No. 2,1e-19;
Matches 118; Conservative 82; Mismatches 216; Indels 99; Gaps 17;

```

```

QY 15 LPSLLSLFLILKRRNRKTRNLPKSGMPFLGTTIGYL--KPYTATLQFMQOH 71
D 15 LAFITSLGIF--FLKQSKSKKFLNLPKPPGMPVYGNLFQVARGKPPF-----EYVDV 67
QY 72 VSKYGIYRSNLFGEPTIVSADAGLNREIIONERLFECSYR----- 114
D 68 RLKYGSIPTLKMGRITMILIDAKLVHEAMIQGATVATRPENPTRIIFSENKFTVNA 127
QY 115 SIGGILGKWSMLVGVDMHRMRSISLNFLSHARLFTILKDVERTHLEFLV-----DSW 168
D 128 TYGPV--WKSL-----RRNWQNMLSSTRKE--FRSVNDNMADKLINRLKDEAE 173
QY 169 QONSIFSAODEAKKFTFNLMKHHMSMDPGEEETEOLKEEVTFTMKGVSA----- 219
D 174 KNGVWVWLKARAVCIILVAMCFGLEMEDEYVERIDQ-----YKSVLITLDRIDYL 229
QY 220 PLNLP--GTAYHKALOSRAITILFERKMEERKLDIKEEDQEEVEKTEDEAEMSKSDH 277
D 230 PLSPEFSQKQKALEVREQVEFLVPIEQRRRAIQPG-----SDHT 273
QY 278 RKQRTDDL-----GWLKHSNLSTEQIIDLISLFFAGHETSSVAIALAIFLQAC 331
D 274 APTFSYLDTLFLDLKVEG--KKSAPSDAELVSLCSEFLNGSTDTTAAVENGIAQLANP 330
QY 332 KAVELEBREHLEIARAKKELGSEELNMDYKMKDFTQCVINETRLGLNVVFLHRKALKD 391
D 331 NVQTKLYEE-----IKRYGEKKVDEKDYKMKYLIANVAVELLRKRPPIHVLTHAVTE 384
QY 392 -VRYKGYDIPSGMKVLPVISAHLDSRYDOPNLFNPMWMOQONNGASSGSGSFSTWGN 450
D 385 PTLGVDYDIPIDANVEYVTPAIADPKWMLNPEKEFDERFISGGEADYTG-----VTGV 439
QY 451 NYMPFGGPRLCASELAKEMAVFIHHLVLKFNW 485
D 440 KMPFVGVRICPGIAMATVAIHILMARMVOEFEM 474

```

```

RESULT 7
US-09-091-432-2
: Sequence 2, Application US/09091432
: Patent No. 5981837
: GENERAL INFORMATION:
: APPLICANT: Chapple, Clint
: TITLE OF INVENTION: A Method For Regulation Of Plant Lignin Composition
: FILE REFERENCE: 7024-325
: CURRENT APPLICATION NUMBER: US/09/091,432
: EARLIER FILING DATE: 1998-06-18
: EARLIER APPLICATION NUMBER: PCT/US96/20094
: EARLIER FILING DATE: 1996-12-19
: EARLIER APPLICATION NUMBER: US 60/009,119
: EARLIER FILING DATE: 1995-12-22
: EARLIER APPLICATION NUMBER: US 60/013,388
: EARLIER FILING DATE: 1996-03-14
: NUMBER OF SEQ ID NOS: 3
: SOFTWARE: Microsoft Word 2.0C
: SEQ ID NO 2
: LENGTH: 520
: TYPE: PRT
: ORGANISM: Artificial
: FEATURE:
: OTHER INFORMATION: Sequence is deduced from DNA sequence of SEQ ID NO:1
: Patent No. 5981837
: US-09-091-432-2

```

```

Query Match          9.8%; Score 264; DB 2; Length 520;
Best Local Similarity 20.3%; Pred. No. 7.6e-18;
Matches 109; Conservative 103; Mismatches 208; Indels 118; Gaps 18;

```

```

: 11 PLLPLSLSLFLILKRRNRKTRFNLPGKSGMPLGERTGYLKPYTATTLGDFMQQ 70
: 15 PTLPLIVVSLFISFTTRRRRP---YPCGRGMPLIGNML-MMDQLTHGGLANL--- 67
: 71 HVSKGKTYRNSLFEPTIVSADAGLNRFILONEGRLE-----CSTPRSTIGIL 120
: 68 -ARKYGGCHLMGFLAHYAVSSPEVAROVLOVDSFSPNRPATIAISLYTRADRAD--- 122
: 121 GKMSLVLVGDMHRMRSIS-LNFLSHARLRTILKDYERHRTFLVLDMSQNSISADE 179
: 123 ---MAFAHYCPWRQMKVVCVKVSRKR-----AESWA---SVRDE 158
: 180 AKKTFENL-----MAKHISM-----DPGEETEOLKKEVTFMKGV 216
: 159 VDKWVRSVSCNVCNRPINVEQIFALTRNITRYARAFGSACENKODEFIRLQEF-SKLGCA 217
: 217 VSAPLNDPGTAY-----HKALQSRATILKFIERKMEERKLDIKEEDQEEVEVTE-DE 268
: 218 FNVADFIYFGVIDPQINKRLVKARNDLDFID-----DIIDENHKKKENONAVDD 269
: 269 AEMSSSDVNRKQRTDDLLGLWVLKHSNLSTE-----QILDILSLLRAGHET 315
: 270 GDVVDTDWV-----DILLAFYSEPAKLVESTADLONSIKITLRNIIKAIMVMVGGET 323
: 316 SSVAIALAIFFOACPKAVEELREELHETARAKELGESELMWDYKMDFOCYINETL 375
: 324 VASALEMALTLLRSPEDLKRQOGLAEVVGIDRVEES-----DIEKITYIKLTKETL 378
: 376 RLGNVVRFLHRRKALKDVRYKGYDIPSGMKVLFVISAVHLDNSRYDOPULFNWRMOQONN 435
: 379 RMHNPRIPLLHETADTIDIGFFIPKKSVMINAFIRGDSPTWDDPFFRSPRLE--- 435
: 436 GASSGSSSFTWGNKNNVMPFGGCPRLCAGSELAKEAMVFTHLVLKFNWELAEEDOP 493
: 436 ---PGVDFKGSNFEETIPFGSGRSCPMGLYALDLAVALHILHCTFWKLDPGMKP 489

```

RESULT 8
US-08-457-274A-25

```

: Sequence 25, Application US/08457274A
: Patent No. 5734086
: GENERAL INFORMATION:
: APPLICANT: Scott, Jeffrey G.
: APPLICANT: Tomita, Takashi
: TITLE OF INVENTION: Cytochrome P450lpr Gene and Its Uses
: NUMBER OF SEQUENCES: 29
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Nixon, Hargrave, Devans & Doyle
: STREET: P.O. Box 1051, Clinton Square
: CITY: Rochester
: STATE: New York
: COUNTRY: USA
: ZIP: 14603
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentIn release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/457,274A
: FILING DATE:
: CLASSIFICATION: 800
: ATTORNEY/AGENT INFORMATION:
: NAME: Goldman, Michael L.
: REGISTRATION NUMBER: 30,727
: REFERENCE/DOCKET NUMBER: 19603/240 (D-1519)
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 716-263-1304
: TELEFAX: 716-263-1600
: INFORMATION FOR SEQ ID NO: 25:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 504 amino acids
: TYPE: amino acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: protein
: HYPOTHEICAL: NO
: ANTI-SENSE: NO
: ORIGINAL SOURCE:
: ORGANISM: Rat
: STRAIN:
: DEVELOPMENTAL STAGE: Adult
: POSITION IN GENOME:
: CHROMOSOME/SEGMENT:
: US-08-457-274A-25

```

```

Query Match          9.8%; Score 263.5; DB 1; Length 504;
Best Local Similarity 24.1%; Pred. No. 8.1e-18;
Matches 132; Conservative 89; Mismatches 213; Indels 113; Gaps 25;

```

```

: 9 LPLPLSLSLFLILKRRNRKTR--FNLDP--PGKSGMPLGERTGYLKPYTATTL 64
: 3 LLSALTLEWVLLAVLVLLYFGGRTGHLFFKKQIGPKRPLPFGTYLVNYT-----MKL 57
: 65 GDFMOQHSKYKGIYSNLF-GE-PTIVSADAGLNRFILONEGRLEEC-----SYPRSIG- 117
: 58 WFEVDECHKKYKGIW--GLFDGOMPLFAITDTMJKNVL-----YECPSVFTNRDRDFP 110
: 118 -GILKMSLVLVGDMHRMRSISLNFLSHARLRTILKDYERHRTFLVLDMSQNS---- 172
: 111 VQIMKRAVSANDEWKRKRALISPTFIS-GRLKE-MEPIIIOYGDILVKYLKQAEETGK 168
: 173 -----IFSA--QDEAKKTFENLMAKHISM-MDPGEETEOLK--EYTFMKGVASAPL 221
: 169 PYTMKKVREAYAMDVITSTSGVNVNDSLNNPKDPVEYKTKKLLRDFPRLPLSVLFPF 228
: 222 NLPGTAYHKAAL-----QSRATILKFIERKMEERKLDIKEEDQEEVEVTEDEAEMSKS 274
: 229 LTP---IYELMNICMPPKDSIEFFKFFVYR--MKETRL----- 261
: 275 DNVNRKQRTDDLLGLWV-----KHSNLSSTEQIIDLILSLLRAGHETSSVAIALAI 324

```

```

Db 262 DSVOKHRV--DFLOLUMMANHDSKDKESTALSDMEITAQSIIFFAGYEPTSTLSFVL 319
QY 325 FFLQACRAVELEEBHELEIARAKKEGESELNDYKKMFOTOCVINETLRIGNVAFRL 384
Db 320 HSLATHPTQKKLOE--EIDRALP--NKAPPTYDTVMEMEYLDVNLTRLPVIGNRL 374
QY 385 HRKALKDVRKYGYIPSGWKVLPVISAHLNDSRYDQPNLFNPMRMOQONNGASSSGSGS 444
Db 375 ERVCKKDVEINGVMPKGSVMIPSYALHRDPQHMPEEERPERFSKENGSIDP----- 430
QY 445 FSTGNNMYPGGGPRCAGSELAKLEMAVFIHHLVLFKNELAEDDQPFAPFVDPENG 504
Db 431 -----YVLPFGNGPNCIGRMFALMNMKLLATKVLQNFSPQCKETO----- 473
QY 505 LPIRVSR 511
Db 474 IPLKLSR 480

```

RESULT 9

PCT-US95-05758-25
Sequence 25, Application PC/TUS9505758

GENERAL INFORMATION:
APPLICANT: Cornell Research Foundation, Inc.
TITLE OF INVENTION: Cytochrome P4501pr Gene and Its
TITLE OF INVENTION: Uses
NUMBER OF SEQUENCES: 29
CORRESPONDENCE ADDRESS:
ADDRESSEE: Nixon, Hargrave, Devans & Doyle
STREET: P.O. Box 1051, Clinton Square
CITY: Rochester
STATE: New York
COUNTRY: USA
ZIP: 14603
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US95/05758

FILING DATE:
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Goldman, Michael L.
REGISTRATION NUMBER: 30,727
REFERENCE/DOCKET NUMBER: 19603/241 (D-1519)
TELECOMMUNICATION INFORMATION:
TELEPHONE: 716-263-1304
TELEFAX: 716-263-1600
INFORMATION FOR SEQ ID NO: 25:
SEQUENCE CHARACTERISTICS:
LENGTH: 504 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHEICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Rat
STRAIN: Unknown
DEVELOPMENTAL STAGE: Adult
POSITION IN GENOME:
CHROMOSOME/SEGMENT: Unknown
PCT-US95-05758-25

Query Match 9.8%; Score 263.5; DB 5; Length 504;
Best Local Similarity 24.18; Pred. No. 8.1e-18;
Matches 132; Conservative 89; Mismatches 213; Indels 113; Gaps 25;

```

QY 9 LPLLLPSSLLSLLFLILKRNRKTR--FNLP--DGKSGMPELGETIGYLKRYTATTL 64
Db 3 LSLATLETFWLLAVLVLLVGLFGTRHGLFKKGIPOGPKPLPFPGVLTNY-----MGL 57
QY 65 GDPMQHVSKTKGKTYRSLNF-GE-PTIVSADAGLNRLFLQNEGLFEC---STPRSTG- 117
Db 58 WKFDVECHKKKYKGIW--GLFDGQMPFLAITDTEMIKVNL-----VKECFSYFTNRROFGP 110
QY 118 -GIIGKMSMLVLVDGMHDMDSISLNFSLHARLRTILKIDVERHTLPVLSMOONS----- 172
Db 111 VLMGKAVSAKDEWKRYRALISPTFTS-GRLE-MPPIEQGDIILVTKLKQDAETGK 168
QY 173 -----IFSA--ODEAKKFTPLMAKHIMS-MDPEEETDOLK--EYVTPMKGVSAPL 221
Db 169 PVTMKKVGAVSMQVYITSTSGVAVDLSLNPCKDPFEVETKLLRPFDPFLSLVLEPF 228
QY 222 NLPGTAHKKAL-----QSRATILKFTERMEERKLDIKEEDQEBEEVKTEDAEKMS 274
Db 229 LTP--IYEMLNICMPKDSIEFFKKFYVR-MKETRL----- 261
QY 275 DHVRKORTDDDLGWL-----KHSNLSSTEQILDLILSLPAGHETSVAIALAI 324
Db 262 DSVOKHRV--DFLOLUMMANHDSKDKESTALSDMEITAQSIIFFAGYEPTSTLSFVL 319
QY 325 FFLQACRAVELEEBHELEIARAKKEGESELNDYKKMFOTOCVINETLRIGNVAFRL 384
Db 320 HSLATHPTQKKLOE--EIDRALP--NKAPPTYDTVMEMEYLDVNLTRLPVIGNRL 374
QY 385 HRKALKDVRKYGYIPSGWKVLPVISAHLNDSRYDQPNLFNPMRMOQONNGASSSGSGS 444
Db 375 ERVCKKDVEINGVMPKGSVMIPSYALHRDPQHMPEEERPERFSKENGSIDP----- 430
QY 445 FSTGNNMYPGGGPRCAGSELAKLEMAVFIHHLVLFKNELAEDDQPFAPFVDPENG 504
Db 431 -----YVLPFGNGPNCIGRMFALMNMKLLATKVLQNFSPQCKETO----- 473
QY 505 LPIRVSR 511
Db 474 IPLKLSR 480

```

RESULT 10

US-08-313-075A-30
Sequence 30, Application US/08313075A
Patent No. 5639870

GENERAL INFORMATION:
APPLICANT: Holton, Timothy A.
APPLICANT: Tanaka, Yoshikazu
TITLE OF INVENTION: GENETIC SEQUENCES ENCODING FLAVONOIDS
TITLE OF INVENTION: PATHWAY ENZYMES AND USES THEREFOR
NUMBER OF SEQUENCES: 58
CORRESPONDENCE ADDRESS:
ADDRESSEE: Scully, Scott, Murphy & Presser
STREET: 400 Garden City Plaza
CITY: Garden City
STATE: New York
COUNTRY: U.S.A.
ZIP: 11530
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/313,075A
FILING DATE: 30-NOV-1994
CLASSIFICATION: 800
PRIOR APPLICATION DATA:
APPLICATION NUMBER: AU PL 1538/92
FILING DATE: 27-MAR-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: AU PL 6698/93

FILING DATE: 07-JAN-1993
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: AU PCT/AU93/00127
 FILING DATE: 25-MAR-1993
 ATTORNEY/AGENT INFORMATION:
 NAME: Digilio, Frank S.
 REGISTRATION NUMBER: 31,346
 REFERENCE/DOCKET NUMBER: 9433
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (516) 742-4343
 TELEFAX: (516) 742-4366
 TELEX: 230 901 SANS UR
 INFORMATION FOR SEQ ID NO: 30:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 476 amino acids
 TYPE: amino acid
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 US-08-313-075A-30

Query Match 9.6%; Score 257; DB 1; Length 476;
 Best Local Similarity 24.0%; Pred. No. 3.3e-17;
 Matches 123; Conservative 87; Mismatches 227; Indels 76; Gaps 21;

12 LLLPSSLILLLFLILKRRNRKTRFNLPPKSGMPFLGRTIYLKPTATTLGDFMQH 71
 9 LLIIVSLISLKKW---KSKNCQTK-KLPBPMKVPFLGSLHMGGLPHHVLRLD--- 59
 72 VSKYGIYRSNLFGEPTIVSADAGLNRFILONEGRLEFCSY-PRSIGILGKMSMLVLV- 129
 60 AKKGPIHNDLGKISAVVTSPEAKRVLTHTDLAF--AKRPILGLEIYCYNSDIAF 117
 130 ---GDHMDMRIS-LNFLSHARLTILKDYERHTLFVLDSSQONS-----IFS 175
 118 SPYGDYWMQMRKICVLEVLSAKNVRS--FNSIRDEILIMIDFLRSSGKPVNITERIFS 175
 176 AQDEAKRTFPLMAKHIMSDPGE-EETDQKKEVTFMKGVASPLPLPGTAV-HKMLQ 233
 176 -----FTSMICKSVGKRIKEDDCIRHVKKWTGLIDGVDADI-FPSLRFLHVLIG 227
 234 SRATILKIERKMERKLDKEEDQEEVEKTEDEAKMSKSDHVKORTDDLLGWLKH 293
 228 MKGRIMD-VHRRKV-----DAIVEVNNHK-ELLRTGKTNGEVEGDELLIVLRL 275
 294 SN-----LSTEDLLIISLLFAGHETSSVAIALAIFLQACPKAVEELREHLEIAR 346
 276 KEEGDQLPITNDNIKAIFNDMFAAGTETSTTIMWAVEILMKNPSVFAKQAQAEVREYK 335
 347 AKKELGESELMWDYKKKDFQOCVINTLRL-GNVVRLHRLKALKDYKGYDIPSGMKV 405
 336 GKEFFDE---DDEELNLYKLVIREFLRHLPRLPLLPRECRHETETINGTTPPLNTRV 390
 406 LPVISAVALHNSRYDQPLFLNPMRMOQONNGASSGSGSFSTGNN--YMPFGGPRICA 463
 391 IVNWALIGRDKYMDADASFPEREH-----NSLMPAGNPFYLPFGSRRICP 440
 464 GSELAKLEMAVFIHHLVLKFNWELAEDDQPAF 496
 441 GISFGIANVYHPLAQLLYHFMRLPPTGVDPNDF 473

RESULT 11
 US-08-201-118-7
 Sequence 7, Application US/08201118
 Patent No. 5786191
 GENERAL INFORMATION:
 APPLICANT: GOLDSTEIN, Joyce A.
 APPLICANT: ROMKES-SPARKS, Marjorie
 TITLE OF INVENTION: CLONING AND EXPRESSION OF COMPLEMENTARY
 TITLE OF INVENTION: DNAs FOR MULTIPLE MEMBERS OF THE HUMAN CYTOCHROME P450 2C
 SUBFAMILY
 NUMBER OF SEQUENCES: 44

CORRESPONDENCE ADDRESS:
 ADDRESSEE: Townsend and Kourie and Crew
 STREET: 379 Lytton Avenue
 CITY: Palo Alto
 STATE: California
 COUNTRY: US
 ZIP: 94301
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patent in Release #1.0, Version #1.25
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/201,118
 FILING DATE: 22-FEB-1994
 CLASSIFICATION: 435
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 07/864,962
 FILING DATE: 09-APR-1992
 ATTORNEY/AGENT INFORMATION:
 NAME: Liedeschuetz, Joe
 REGISTRATION NUMBER: 37,505
 REFERENCE/DOCKET NUMBER: 15280-192-1
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (415) 326-2400
 TELEFAX: (415) 326-2422
 INFORMATION FOR SEQ ID NO: 7:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 490 amino acids
 TYPE: amino acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 US-08-201-118-7

Query Match 9.6%; Score 256.5; DB 1; Length 490;
 Best Local Similarity 20.8%; Pred. No. 3.9e-17;
 Matches 110; Conservative 99; Mismatches 202; Indels 119; Gaps 20;

11 LLLPSSLILLLFLILKRRNRKTRFNLPPKSGMPFLGRTI-----GYLKYATATLG 65
 3 PFVVLVLSTFLFLSLMRQSCRK--LPQPTPLPIIGNMLQIDVDICKSF----- 55
 66 DPMQVSK-YGIYRSNLFGEPTIVSADAGLNRFILONEGRLEFCSYPRSIGILGK-W 123
 56 -----NFSKVVGPVTVFEGMNPVIVFHYGEAVKDALIDNGEEFSGRGNPSISQRTKGL 110
 124 SMLVLVGMHMDMRISLNFLSHARLTILKDYERHTLFVLDSSQONS-----FVL 165
 111 GIISNGRMRKEIRRFSLNLFNFGMKRSIEDRYOEAHCLVEELKRTKASPCDPTFL 170
 166 DSMQONSIFSADEAKKRTFNLMAKHIMSDPGEETEOLKKEVTFMK-----GVSA 219
 171 GCAPCNVICSVYQ-KRVDYK-----DQNFLMKRFENENRIINS 210
 220 PL-----NLPGTAHKAQSRATILKTERKMEERKLDKEEDQEEB-----V 263
 211 PWIQYCNPPRLIDCFPGT-HNKVLKNVALTRSYIRKVEKHQASLDVNNPDMDFMCEFL 269
 264 KTEDEAKMSKDHVAKKQRTDDLLGWLKHSNLSREQLIDLLSLFLPGHETSSVAIALA 323
 270 KWEDEKDKQKSE-----FNTEMLVGTAVDLVAGTETSTTLRYG 309
 324 IFFLQACPKAVEELREE--HLEIARAKKELGSELNMDYKKMDTQCVINTETRLGNVY 381
 310 LLLLKHPREVAKVQEBIDNV-IGRHSRSC-----MDRSMPRTDAVNHITQYSDLY 362
 382 -RFLHRLKALKDYRYKGYDIPSGMKVLPVISAVALHNSRYDQPLFLNPMRMOQONNGASS 440
 363 PTGVPHAATTDTKFRNYILPKGTTIMALLTSVLHDKKEPNNPNDIPDHFIDKN----- 416
 441 GSGSFTGWNNNMPFGGGRICAGSELAKLEMAVFIHHLVLKFNWELAE 490

Db 417 --GNFKK-SDYFMPFSAGKRICAGEGLARMELFLFTIITLONFNLSVD 463

RESULT 12

US-08-238-821B-7
Sequence 7, Application US/08238821B
Patent No. 5912120
GENERAL INFORMATION:
APPLICANT: GOLDSTEIN, Joyce A.
APPLICANT: ROMKES-SPARKS, Marjorie
TITLE OF INVENTION: CLONING, EXPRESSION AND DIAGNOSIS OF HUMAN
TITLE OF INVENTION: CYTOCHROME P450 2C19: THE PRINCIPAL DETERMINANT OF S-
NUMBER OF SEQUENCES: 61
CORRESPONDENCE ADDRESS:
ADDRESSEE: Townsend and Townsend and Crew LLP
STREET: Two Embarcadero Center, 8th Floor
CITY: San Francisco
STATE: California
COUNTRY: US
ZIP: 94111
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/238,821B
FILING DATE: 06-MAY-1994
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/201,118
FILING DATE: 22-FEB-1994
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/864,962
FILING DATE: 09-APR-1992
ATTORNEY/AGENT INFORMATION:
NAME: Liebeschuetz, Joe
REGISTRATION NUMBER: 37,505
REFERENCE/DOCKET NUMBER: 15280-192110US
TELECOMMUNICATION INFORMATION:
TELEPHONE: (650) 326-2400
TELEFAX: (650) 326-2422
INFORMATION FOR SEQ ID NO: 7:
SEQUENCE CHARACTERISTICS:
LENGTH: 490 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-238-821B-7

Query Match 9.6%; Score 256.5; DB 2: Length 490;
Best Local Similarity 20.8%; Pred. No. 3,9e-17;
Matches 110; Conservative 99; Mismatches 202; Indels 119; Gaps 20;

Db 11 PLLLPSSLLFLFLILKRRNRKTRNLPCKSGMPFLGETI-----GYLKPYATTATLG 65
Db 3 PFVAVLCLSPMLFLSLMRGCRRK--LPPGPTPLPIIGMLDIDVKICKST----- 55
Qy 66 DFMQOHVSK-IGKTYRNSLGEPTIVSADAGLNFITLONSGRLFECSYPSISGILGK-W 123
Db 56 -----NESKYVGPVETVYFGNNPIVPHGVEAVKALIDNGEERSGKNSPISQITKGL 110
Qy 124 SMLVYGDMDHMDSISLNFSLSHARLRTIILKDD---VERHTL-----FVL 165
Db 111 GIISSNKRKRKEIRFSTILNIRNFGMKRSIEDRYQEAHCLVLEELKRTKASPCDPPFIL 170
Qy 166 DSMOONSIFSAODEAKRTFTLMAKHIMSDPGEEETEOLKKEYVTMK-----GVVSA 219

Db 171 GCAPCNVICSVFO-KREDDYK-----DQNFLETLMKRENNENRILNS 210
Qy 220 PL-----NLPTAHKALQSRATILKFIERKMEERKLDIKEEDDEEE-----V 263
Db 211 FMIQVGNPFLLIDCFPT-HNKVKLVNLTSTRYIREKVEHQASLDVNNRDRPMDCEFLI 269
Qy 264 KTEDEAEKSKSDHYRKQRTDDLLGWVLKHSNLSSTEOILDLILSLFAGHETSSVAIALA 323
Db 270 KMEDEKDNQKE-----FNENLVGVADLFAAGTTTSTTLRYG 309
Qy 324 IFELQACPKAVEELREE--HLEIARAKKEIGESSELMDYKMKDFTQCVINETLRGNV 381
Db 310 LLLILKHEVYAKVQOEIDHV-IGRHRSPC-----MODRSHMPTDAVVAHEIORSDLV 362
Qy 382 -RFLHRAKLVKRYKGYDIPSGWKVLPYISAVHLDNSRYDDPNLEFNPMMQOONNGASS 440
Db 363 PTGVPNAVTTDKFRNRYLIPGTTIMALLTSVLDDKEFPNPIFDQHPFLDKN----- 416
Qy 441 GSGSFTHWNNYMPFGGPRLCAGSELAKLEMAVFIHHLVLKPFMELAED 490
Db 417 --GNFKK-SDYFMPFSAGKRICAGEGLARMELFLFTIITLONFNLSVD 463

RESULT 13

PCT-US95-05744-7
Sequence 7, Application PC/TUS9505744
GENERAL INFORMATION:
APPLICANT: GOLDSTEIN, Joyce A.
APPLICANT: ROMKES-SPARKS, Marjorie
TITLE OF INVENTION: CLONING, EXPRESSION AND DIAGNOSIS OF HUMAN
TITLE OF INVENTION: CYTOCHROME P450 2C19: THE PRINCIPAL DETERMINANT
NUMBER OF SEQUENCES: 61
CORRESPONDENCE ADDRESS:
ADDRESSEE: Townsend and Townsend and Crew
STREET: 379 Lytton Avenue
CITY: Palo Alto
STATE: California
COUNTRY: US
ZIP: 94301
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US95/05744
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/238,821
FILING DATE: 06-MAY-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/201,118
FILING DATE: 22-FEB-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/864,962
FILING DATE: 09-APR-1992
ATTORNEY/AGENT INFORMATION:
NAME: Dow, Karen B.
REGISTRATION NUMBER: 29,684
REFERENCE/DOCKET NUMBER: 15280-192-1-1
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 326-2400
TELEFAX: (415) 326-2422
INFORMATION FOR SEQ ID NO: 7:
SEQUENCE CHARACTERISTICS:
LENGTH: 490 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear

MOLECULE TYPE: protein
PCT-US95-05744-7

Query Match 9.6%; Score 256.5; DB 5; Length 490;
Best Local Similarity 20.8%; Pred. No. 3.9e-17;
Matches 110; Conservative 99; Mismatches 202; Indels 119; Gaps 20;

```

QY 11 LLLLLPILLSLLPILLLKRRNRKTRFNLPRGKSGMPFLGETI-----GYLKPYATTLG 65
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 3 PFVVLVLCLSLFLLSLMRQSCRK--LPPGPPPLPIIGMLQIDVKICKSFT----- 55
QY 66 DPMOHWK-VGKIYRSLFGEPTIVSADAGLNRFILONEGRLEFCSPRSIGLIGK-W 123
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 56 -----NFKYVGPVTVTFGNAPIVFHYGEAVKALIDNGEFGSGRNSPISOKITGGL 110
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 124 SMLVLVGMHMDRSLNPLSHARLRTLLKD---VERHTL-----EVL 165
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 111 GIISSNGKRMKEIRRFSLTNLRNFCMKRSIEDRYOEBAHCLVEELRKTAKSPCDPTFL 170
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 166 DSMOONSIFSAODEKKTFTNLMAKHIMSMDGEEETQLKEYTFMK-----GVYSA 219
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 171 GCAPCNVICSVVFQ--KRPDYK-----DONFLTLKRFENENFRILNS 210
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 220 PL-----NLPGTAHKALOSRATILKFTERKMEERKLDIKEEDQEEB-----V 263
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 211 PMIQVCNPFLLIDCFPGT-HNKVLKANALTRSYIREKVKHQASLDVNNRPDPCFLI 269
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 264 KTEDEAKSKSDHVRKORTDDLLGVLKHSNLSLQIIDLILSLFAGHETSVVALA 323
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 270 KMEDEKDKQKSE-----FNLENLVGVADLFVAGTETSTTLKRG 309
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 324 IFFLQACRAVELEEE--HLEIARAKKELGESLNMDDYKMDTQCVINTELKGVV 381
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 310 LLLLKLBVEVTAQVQELDHW-IGHRSPC-----MODRSHMPTDVAVHDIQRYSDLV 362
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 382 -RFLRKALKDVRKGYDIPSGMKVLPVISAHLNDSRYDQPNLEPNRMQOONNGASSS 440
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 363 PTGVHAHTTPTDKFRNYLIPKGTITMALLTSVLDHDKKEFPNPNITDPGHFLDK----- 416
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 441 GSGSSTWGNNTMPGGPRLCAGSELAKLEMAVFIHHLVLKFNWELAD 490
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 417 --GNFKK-SDYFMPFSACKRICAGEGLARMELELFTITLQNFNLKSVDD 463
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

RESULT 14
US-08-532-065B-2
Sequence 2, Application US/08532065B
Patent No. 5753507

GENERAL INFORMATION:
APPLICANT: Ohta, Daisaku
APPLICANT: Mizutani, Masaharu
TITLE OF INVENTION: Plant Geraniol/Nerol 10-Hydroxylase and
NUMBER OF SEQUENCES: 6
CORRESPONDENCE ADDRESS:
ADDRESS: No. 5753507artis Corporation
STREET: 59 Route 10
CITY: East Hanover
STATE: NJ
COUNTRY: USA
ZIP: 07936
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/532,065B
FILING DATE: 22-SEP-1995
CLASSIFICATION: 800
ATTORNEY/AGENT INFORMATION:
NAME: Meigs, J. Timothy

REGISTRATION NUMBER: 38,241
TELECOMMUNICATION INFORMATION:
TELEPHONE: 919-541-8587
TELEFAX: 919-541-8689
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 495 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-532-065B-2

Query Match 9.1%; Score 245; DB 16; Length 495;
Best Local Similarity 22.9%; Pred. No. 5.6e-16;
Matches 120; Conservative 80; Mismatches 218; Indels 106; Gaps 19;

```

QY 12 LLLPILLS-LLLPILLLKRRNRKTRFNLPRGKSGMPFLGETIGYLYKPYATTLGDFMQQ 70
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 10 LLLPFIILSCPIFTTTSGRISRGATALPPGPPPLPIIG-NIHLVCKNPHRSFALSK- 67
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 71 HYSKYGKIYRSLFGEPTIVSADAGLNRFILONEGRLEFCSTP-----RSIG----- 117
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 68 --TYGPVMSLKLSGLNTVIVIASPEAREVLTHTQILSARSPNAVRISINHQDASLVWL 124
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 118 -GIIGKSMVLVGMHMDRSL-NFLSHARL--RTILKDVERTLPLVDSMOONS 172
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 125 PSSSARWRLL-----RRLSVTQLSPQRIATATLARNKKVKELVSFSSSDREE 174
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 173 IFSADDEAKKFTFNMAKHIMSMDGEEETQLKEYTFMKGVVSADNLPGT----- 226
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 175 SVDISRVAFTITLNIISNLSFSDLSYNA---KASINGVDYIVISWDAAGTDAANY 230
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 227 ---AVHKALOSRAITLK-----FLRKMEERKLDIKEEDQEEBVEYKTEDEAM 271
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 231 PFPLRFLDLQGVKTKFCTERLVVFRFGFTDAKIAEKSSQNNPMD-----V 277
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 272 SKSDHVRKORTDDLLGWLKHSNLSLQIIDLILSLFAGHETSVVALAIFLQAC 331
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 278 SKNDFV-----DNLIDYKGESELSIDIEHLIDMTAGTDTSSLTLEWPTLNLNP 331
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 332 KAVEELREHELEIARAKE---LGESL-NMDDYKMDTQCVINTELKGVVRL-H 385
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 332 KT-----MAKAQAEIDCVIQNGVIEESDISKLPYLQAVVKEFRFLHTPPVLLIP 381
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 386 RKALDVRKGYDIPSGMKVLPVISAHLNDSRYDQPNLEPNRMQOONNGASSSGSGSF 445
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 382 RKAESDAEILGFMVLKDQVLVNWVAIGRDPVMDNPQEFEBERFLGKD-----M 431
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 446 STWGNNTMPGGPRLCAGSELAKLEMAVFIHHLVLKFNWEL 487
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 432 DVGRDVEITLPPGAGRICPGMPLMAKTIVSLMLASLVSFDWKL 475
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

RESULT 15
US-08-948-564-16
Sequence 16, Application US/08948564
Patent No. 6121512

GENERAL INFORMATION:
APPLICANT: Saminszky, Balazs
APPLICANT: Dewey, Ralph E.
APPLICANT: Corbin, Frederick T.
TITLE OF INVENTION: No. 6121512el Cytochrome P-450 Constructs and
NUMBER OF SEQUENCES: 23
CORRESPONDENCE ADDRESS:
ADDRESS: Virginia C. Bennett
STREET: PO Box 37428
CITY: Raleigh
STATE: NC
COUNTRY: USA
ZIP: 27627
COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/948,564
FILING DATE:
CLASSIFICATION: 800
ATTORNEY/AGENT INFORMATION:
NAME: Bennett, Virginia C.
REGISTRATION NUMBER: 37,092
REFERENCE/DOCKET NUMBER: 5051-409
TELECOMMUNICATION INFORMATION:
TELEPHONE: 919-854-1400
TELEFAX: 919-854-1401
INFORMATION FOR SEQ ID NO: 16:
SEQUENCE CHARACTERISTICS:
LENGTH: 576 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-948-564-16

Query Match 9.1%; Score 244.5; DB 3; Length 576;
Best Local Similarity 22.7%; Pred. No. 7.9e-16;
Matches 111; Conservative 91; Mismatches 181; Indels 107; Gaps 22;

QY 63 TLGDMQOHVSKYKITYSNLEGEPTIVSADAGLNFILO-----NGRLFECSYPRSI 116
DB 97 SLYDMFLFH-----GAVYKLAGPKAFVVSPIYARHILRENAFSYDKGLADILEP--- 149
QY 117 GGIGCK-----WSM--LVIVGDMHRMRSISLNFSLHARLTLL--KDERHHTLF 163
DB 150 --IMKGLIPADLTWKQRRIYAFNFIHNSYLEAMVKIFTTCSERTILKFNKLLLEGECYD 207
QY 164 VLDWMQONSIFSADDEAKKFTENLMAKHIMSDPGE--EETEOLKEEYVTFMGVVSAPL 221
DB 208 GPDSIE---LDLEAEFSSALDITGLGVFNVDGFSVTKESPVIKAVYGTLEAEHRSTF 263
QY 222 NLPGTAHYKALQSRTILKFTERRKKEERKLDIK-----EEDQEEEEKYKTEDE 268
DB 264 YIP--YKRIPLARVIYPR--QRKFQD--DLKVINTCGLDGLIRNAKESROETDVE--- 311
QY 269 AEMSKSDHVRKQRTDDDLGWN--LKHNSLSTEQIIDLITLSLFGHETSVAIALAIF 326
DB 312 -KLOQRDYLNK--DASLLRFLVDMRGADVDDRLDMLMTNLINAGHETTAVALTWAVFL 368
QY 327 LQACPKAVEELREHLEIETARRAKKEIGESELNWDYKKMDFTCVINEETLRIGNVVRFLLR 386
DB 369 LAQNPSSKKRAQ-----AEVDLVLTGTGRPTFESEIKELQYIRLIYVEALRLYPOPLLIR 422
QY 387 KALK-DV-----RYKGYDIPSGKVLPIYSAVHLNDSRY--DQPNLFNPMRMOQONG 436
DB 423 RSLKSDVLPDGGHKGKGDYALPAGTDF--ISVYNLHRSPLYFMDRPPDDFEPRFLYQKN 480
QY 437 ASSSGSGSFTWGN-----NYMPGGGPRLCAGSELAKLEMAVFIH 477
DB 481 EIEIG-----WAGIDPSPRSGLALYPNVISDFALFLPRGGSPKCVGDQFALMESTVALT 534
QY 478 HLVLKENNEL 487
DB 535 MLQNFDEL 544

Search completed: October 17, 2001, 22:07:08
Job time: 10067 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 17, 2001, 19:51:16 ; Search time 41.56 Seconds
(without alignments)
940.268 Million cell updates/sec

Title: US-09-502-426-2

Perfect score: 2681

Sequence: 1 MFETENHTLLPILLPLSLLS.....FAFPVDFPNGLPVRSRL 513

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 219241 seqs, 76174552 residues

Total number of hits satisfying chosen parameters: 219241

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

PIR_68:**
1: PIR1:**
2: PIR2:**
3: PIR3:**
4: PIR4:**

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description |
|------------|-------|-------------|--------|----------|--------------------|
| 1 | 2677 | 99.9 | 513 | 2 T46143 | steroid 22-alpha-h |
| 2 | 978.5 | 36.5 | 472 | 1 S55379 | cytochrome P450 cy |
| 3 | 866 | 32.3 | 512 | 2 H96759 | probable steroid 2 |
| 4 | 860 | 32.1 | 457 | 2 D85429 | cytochrome P450 1i |
| 5 | 815 | 30.4 | 382 | 2 T48613 | hypothetical prote |
| 6 | 761.5 | 28.4 | 464 | 2 T07859 | cytochrome P450 ho |
| 7 | 632 | 23.6 | 457 | 2 T04444 | cytochrome P450 - |
| 8 | 611 | 22.8 | 482 | 2 T02739 | probable cytochrom |
| 9 | 581 | 21.7 | 487 | 1 C71417 | cytochrome P450 d1 |
| 10 | 579.5 | 21.6 | 485 | 2 A84859 | probable cytochrom |
| 11 | 579.5 | 21.6 | 490 | 2 H86185 | hypothetical prote |
| 12 | 577 | 21.5 | 455 | 2 T48973 | cytochrome P450-1i |
| 13 | 547.5 | 20.4 | 489 | 2 B84733 | probable cytochrom |
| 14 | 497.5 | 18.6 | 460 | 2 D96813 | hypothetical prote |
| 15 | 491.5 | 18.3 | 519 | 1 T02263 | cytochrome P450 DW |
| 16 | 489.5 | 18.3 | 255 | 2 T04602 | cytochrome P450 ho |
| 17 | 479.5 | 17.9 | 444 | 1 S75761 | cytochrome P450 - |
| 18 | 369 | 13.8 | 349 | 2 A86329 | hypothetical prote |
| 19 | 357 | 13.3 | 518 | 2 T20908 | hypothetical prote |
| 20 | 338 | 12.6 | 453 | 2 C83722 | cytochrome P450 hy |
| 21 | 329.5 | 12.3 | 517 | 2 T20907 | hypothetical prote |
| 22 | 324.5 | 12.1 | 520 | 2 T24778 | hypothetical prote |
| 23 | 321.5 | 12.0 | 518 | 2 T24783 | hypothetical prote |
| 24 | 316 | 11.8 | 500 | 2 T04737 | cytochrome P450 ho |
| 25 | 310.5 | 11.6 | 520 | 2 T24777 | hypothetical prote |
| 26 | 310 | 11.6 | 500 | 2 T52175 | cytochrome P450 mo |
| 27 | 306.5 | 11.4 | 504 | 2 A25222 | cytochrome P450 3A |
| 28 | 302.5 | 11.3 | 491 | 2 S31277 | cytochrome P450 2B |
| 29 | 300 | 11.2 | 491 | 2 T84735 | testosterone 16a-h |

| | | | | | |
|----|-------|------|-----|----------|--------------------|
| 30 | 299.5 | 11.2 | 492 | 2 S27160 | cytochrome P450 2B |
| 31 | 297 | 11.1 | 491 | 2 A31047 | testosterone 16alp |
| 32 | 297 | 11.1 | 494 | 2 A33293 | cytochrome P450 2A |
| 33 | 294.5 | 11.0 | 491 | 1 O4RBP | cytochrome P450 2B |
| 34 | 294 | 11.0 | 506 | 2 D96672 | probable Cytochrom |
| 35 | 294 | 11.0 | 516 | 2 T48140 | flavonoid 3',5'-hy |
| 36 | 292 | 10.9 | 491 | 2 O4RBP | cytochrome P450 2B |
| 37 | 291.5 | 10.9 | 491 | 2 S35666 | cytochrome P450 2B |
| 38 | 291.5 | 10.9 | 530 | 1 A27491 | lanosterol 14alpha |
| 39 | 290.5 | 10.8 | 502 | 2 T05246 | cytochrome P450 mo |
| 40 | 288.5 | 10.8 | 504 | 2 A29410 | cytochrome P450, g |
| 41 | 287.5 | 10.7 | 520 | 2 T24780 | hypothetical prote |
| 42 | 287 | 10.7 | 503 | 2 JC4702 | cytochrome P450 3A |
| 43 | 284.5 | 10.6 | 505 | 1 S38534 | cytochrome P450 76 |
| 44 | 284 | 10.6 | 501 | 2 T04735 | cytochrome P450 ho |
| 45 | 282 | 10.5 | 491 | 2 T49625 | testosterone 16a-h |

ALIGNMENTS

```

RESULT 1
T46143
steroid 22-alpha-hydroxylase (DWF4) - Arabidopsis thaliana
N:Alternate names: protein T3A5_40
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 04-Feb-2000 #sequence_revision 04-Feb-2000 #text_change 16-Feb-2001
C:Accession: T46143
R:Blöcker, H.; Mewes, H.W.; Mayer, K.F.X.; Lemcke, K.; Schueller, C.; Quetier, F.; S
submitted to the Protein Sequence Database, December 1999
A:Reference number: Z23024
A:Accession: T46143
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-513 <BLO>
A:Cross-references: EMBL:AL132979
A:Experimental source: cultivar Columbia; BAC clone T3A5
C:Genetics:
A:Map position: 3
A:Introns: 74/2; 182/3; 233/3; 338/3; 369/3; 396/1; 432/3
A:Note: T3A5_40
C:Superfamily: Synecocystis cytochrome P450 slr0574; cytochrome P450 homology
F:308-484/Domain: cytochrome P450 homology <P45>
F:462/Binding site: heme iron (cys) (axial ligand) #status predicted

Query Match          99.9%; Score 2677; DB 2; Length 513;
Best Local Similarity 99.8%; Pred. No. 9.2e-166;
Matches 512; Conservative 1; Mismatches 0; Indels 0; Gaps 0.

OY 1 MFETENHTLLPILLPLSLLSLLFLTLKRRNRKTRFNLPPKSGWPLGFTIGYLKPYT 60
    |||
DB 1 MFETENHTLLPILLPLSLLSLLFLTLKRRNRKTRFNLPPKSGWPLGFTIGYLKPYT 60
    |||
OY 61 ATTLDPMQOHTSKGKTYRSLRGEPTIVSADAGLNRIILONEGRIFPCSPRSIGTL 120
    |||
DB 61 ATTLDPMQOHTSKGKTYRSLRGEPTIVSADAGLNRIILONEGRIFPCSPRSIGTL 120
    |||
OY 121 GKSMILVLVGDHMRDRSISLNFSLSHARLRTLLKDVREHTLFVLDSMOONSIFSNODEA 180
    |||
DB 121 GKSMILVLVGDHMRDRSISLNFSLSHARLRTLLKDVREHTLFVLDSMOONSIFSNODEA 180
    |||
OY 181 KFTFNLMARKHMSMDPGEETEOLKKEVVTMKGVSAPLMPGTAYHKAQSRATYIK 240
    |||
DB 181 KFTFNLMARKHMSMDPGEETEOLKKEVVTMKGVSAPLMPGTAYHKAQSRATYIK 240
    |||
OY 241 FTERMEERKRLDKREDDEEVEKTEDEAEKMSQHVRRQRTDDDLGVLKHSNLSTEQ 300
    |||
DB 241 FTERMEERKRLDKREDDEEVEKTEDEAEKMSQHVRRQRTDDDLGVLKHSNLSTEQ 300
    |||
OY 301 IIDLILSLFAGHETSVAIALAIFLQACPRAVELREHLEIARAKKEGESEINMD 360
    |||

```

Db 301 ILDLISLFLRGHETSSVAIALAIFLQACPAVEELREEHLEIARAKKESELNMD 360
 QY 361 YKKMDFTCVINETLRLGNVVRFLRKALKDVRKYGYDIPSGMKVLPYISAVHLDNSRYD 420
 Db 361 YKKMDFTCVINETLRLGNVVRFLRKALKDVRKYGYDIPSGMKVLPYISAVHLDNSRYD 420
 QY 421 QPNLFNPRMOQONNGASSSGSGSFSTWGNVMPFGGPRLCAGSELAKLEMAVFIHLV 480
 Db 421 QPNLFNPRMOQONNGASSSGSGSFSTWGNVMPFGGPRLCAGSELAKLEMAVFIHLV 480
 QY 481 LKFNWELAEEDDOPFAFPVDEPNGLPIRVSR 513
 Db 481 LKFNWELAEEDDOPFAFPVDEPNGLPIRVSR 513

RESULT 2

S55379
 cytochrome P450 CYP90 - Arabidopsis thaliana
 N:Contains: oxidoreductase (EC 1.-.-.-)
 C:Species: Arabidopsis thaliana (mouse-ear cress)
 C:Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 21-Jul-2000
 C:Accession: S55379
 R:Sequences: M.: Nemeth, K.; Koncz, Z.; Nagy, F.; Koncz, C.
 submitted to the EMBL Data Library, May 1995
 A:Reference number: S55379
 A:Accession: S55379
 A:Molecule type: mRNA
 A:Residues: 1-472 <S2E>
 A:Cross-references: EMBL:X87367; NID:9853718; PIDN:CAA60793.1; PID:9853719
 C:Genetics:
 A:Gene: CYP90
 C:Superfamily: Synchocystis cytochrome P450 slr0574; cytochrome P450 homology
 C:Keywords: chromoprotein; electron transfer; heme; iron; metalloprotein; monooxygenase;
 F:275-440/Domain: cytochrome P450 homology <P45>
 F:418/Binding site: heme iron (Cys) (axial ligand) #status predicted

Query Match 36.5%; Score 978.5; DB 1; Length 472;
 Best Local Similarity 41.9%; Pred. No. 5.9e-56;
 Matches 210; Conservative 79; Mismatches 171; Indels 41; Gaps 8;

QY 12 LLLPSLISLFLILKRRNRKTRFNLPRPGSGMPLFGETIGYLYKYATATLGDPMOCH 71
 Db 7 LLLSSIAAGFL--LLLRTRRYRMGLPGSLGLPLIGTFOLIGAYKTENPEPFIDER 63
 QY 72 VSKYGIYRSNLFGPPTVSADAGINRFLTONEGRLEPCSTPRSTGIGLKMSMLVLCGD 131
 Db 64 VARGSVYMTLFGPPTVSADAGINRFLTONEGRLEPCSTPRSTGIGLKMSMLVLCGD 123
 QY 132 MHRMRSISLNLFLSHARLTILKDVRRHTLFLVDSMOQNSIFSNODEAKKFTENLMAKH 191
 Db 124 LHKRHSITNSFANSSITKDLMLDIDRLVFNLDSSSVLL--MEKAKITTELYTKQ 181
 QY 192 IMSDPRGEELQKKVYFMKGVASAPLNLPGTAHYKALQSRTATIKTERKMEERKL 251
 Db 182 LMSDPG--EMSESRKEVLTVEGFSLPLPLFSTYRKALQAR-----RKVAEALT 232
 QY 252 DIKEEDDEEVEKTEDEAKMSDHYVRKQRTDDLLGVTLKHSNLSTEQIIDLISLFLA 311
 Db 233 VVVKRRREEEGAE-----RKMDLALLAADGSDDEIVDFVALLYA 278
 QY 312 GHETSSVAIALAIFLQACPAVEELREEHLEIARAKKESELNMDYKKMDFTCV 371
 Db 279 GYETSTIMTLAVFLFETPLALQKKEHEKIRAKSD--SYSEMSDYKSMFTQCV 336
 QY 372 NETLRGLNVAVFLRKALKDVRKYGYDIPSGMKVLPYISAVHLDNSRYDQPNLFNPRMO 431
 Db 337 NETLRVANIIGVFRRAMTDEIVKIKYKIPKGMVFSFRVAVLDRPHFKDARTFNPRMO 396
 QY 432 QONNGASSSGSFSTWGNVMPFGGPRLCAGSELAKLEMAVFIHLVLFKFWELAE 490
 Db 397 -----SNSVTGSHNVFTFPGGPRLCAGSELAKLEMAVFIHLVLFKFWELAE 446

QY 491 DOPFAFPVDEPNGLPIRVSR 511
 Db 447 DKLVEFPTTRQKRYPIVKR 467

RESULT 3

H96759
 Probable steroid 22-alpha-hydroxylase T9L24.44 [Imported] - Arabidopsis thaliana
 C:Species: Arabidopsis thaliana (mouse-ear cress)
 C:Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 31-Mar-2001
 C:Accession: H96759
 R:Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alon
 Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar,
 ansen, N.F.; Hughes, B.; Huizart, L.
 Nature 408, 816-820, 2000
 A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim,
 C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maitl, R.; Marzla
 Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
 A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallo
 ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
 A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
 A:Reference number: A86141; MUID:21016719
 A:Accession: H96759
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-512 <STO>
 A:Cross-references: GB:AE005173; NID:911120803; PIDN:AAG30983.1; GSPDB:GN00141
 C:Genetics:
 A:Gene: T9L24.44
 A:Map position: 1

Query Match 33.3%; Score 866; DB 2; Length 512;
 Best Local Similarity 33.8%; Pred. No. 1.2e-48;
 Matches 181; Conservative 103; Mismatches 173; Indels 76; Gaps 7;

QY 12 LLLPSLISLFLILKRRNRKTRFNLPRPGSGMPLFGETIGYLYKYATATLGD 66
 Db 11 LLSVSTFLFATITLFLAGIARAKRRAPRLPRGSGMPLDIGNFMAUNAVGSHSS 70
 QY 67 FMOQHVSK-----YKIRSNLFGPPTVSADAGINRFL 101
 Db 71 FVEKQIKFVSLCSVLLILKRPDNGFNIRYGRIFSCSLFGWAVASADPPFNRFIM 130
 QY 102 ONEGRLEPCSTPRSTGIGLKMSMLVLCGDHMRDMSISLNLFLSHARLTILKDVRRHT 161
 Db 131 ONEGRLEPCSTPRSTGIGLKMSMLVLCGDHMRDMSISLNLFLSHARLTILKDVRRHT 161
 QY 162 LFLVDSMOQNSIFSNODEAKKFTENLMAKHIMSMDPGEEELQKKVYFMKGVASAPL 221
 Db 191 LQTLISNFDGEVLLDLCRKVAIHLWVNLGVS--SESEVDMSQLFSDVDGLSVPI 249
 QY 222 NLPGTAHYKALQSRTATIKTERKMEERKLDIKEEDDEEVEKTEDEAKMSDHYVRKQ 281
 Db 250 DLPGFTYKAMKAKKEITIRKINKTIEKRLQKASD-----T 286
 QY 282 TDDLLGVTLKHSNLSTEQIIDLISLFLAGHETSSVAIALAIFLQACPAVEELREEH 341
 Db 287 AGNGVGLKEEESLPNSMDFTIINLFGNETTSKTMFLAVFLTHCPRAMTQLLEEH 346
 QY 342 LEIARAKKESELNMDYKKMDFTCVINETLRLGNVVRFLRKALKDVRKYGYDIPSG 401
 Db 347 -----DRLAGMLTGMQDYKTMDFTCVIDETLRGLGIALWLRKAEKEDVSXYDYVIRK 399
 QY 402 GWKVLPIYSAVHLDNSRYDQPNLFNPRMW-----QONNGASSSGSGSFSTWGNVMPFG 456
 Db 400 GCFVVPPLSAVHDEESYKESLSFNPRMWDLPETQOKRMWTSP-----FYCPFG 449
 QY 457 GGRPLCAGSELAKLEMAVFIHLVLFKFWELAEEDDOPFAFPVDEPNGLPIRVSR 511
 Db 450 GGRPLCAGSELAKLEMAVFIHLVLFKFWELAEEDDOPFAFPVDEPNGLPIRVSR 504

```

RESULT 4
D85429
cytochrome P450 like protein [imported] - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cross)
C:Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 02-Apr-2001
C:Accession: D85429
R:Anonymous, The European Union Arabidopsis Genome Sequencing Consortium, The Cold Spring
Nature 402, 769-777, 1999
A:Title: Sequence and analysis of chromosome 4 of the plant Arabidopsis thaliana.
A:Reference number: A85001; MUID:20083488
A:Accession: D85429
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-457 <STO>
A:Cross-references: GB:NC_001268; NID:g2720586; PIDN:CAB80304.1; GSPDB:GN00140
C:Genetics:
A:Gene: AT4g36380
A:Map position: 4
C:Superfamily: Synechocystis cytochrome P450 slr0574; cytochrome P450 homology
C:Keywords: heme; iron; metalloprotein
F:396/Binding site: heme iron (Cys) (axial ligand) #status predicted

Query Match 32.1%; Score 860; DB 2; Length 457;
Best Local Similarity 36.8%; Pred. No. 2,66-48;
Matches 179; Conservative 93; Mismatches 160; Indels 54; Gaps 8;

OY 39 LPPKSGPFLGETIGYL-----KPYATTLGDFMOOHVSKYKITYSNLFGEPTIVSA 92
D 1 PINGSLPVPVIGETLNFACGSSRPVT-----FMDRKSLYGVKFTNIIITPFIIST 55
OY 93 DAGLRFTLQDNGRLFECSYPSRISGILGKMSMLVLDGMDHMDMSISLNFISHARLTI 152
D 56 DAENKAVVLQNHGNTFVATPKSTIELGENSELISINCPHQRHLTLGAFURSPHLDR 115
OY 153 LKDYERHTLEVLDSMOONSIFSAODEAKKFTFNLMARKHNSMDGEETEOLAKREYTF 212
D 116 ITRDIEASVVLTLASMAQLPLVHODEIKMTFELVVKLSTSG-EDMNLKLEFEFF 174
OY 213 MKGVSAALNLPRTAYHAKALOSRATILKFERKMEERKLDIKEEDQEEBEVKT-----E 266
D 175 IKGILCIPIKPFGRFLYKSLAKERLIKMKVKKVVERQVAMTTSPADVDVLLRSGD 234
OY 267 DEAEKSDHYRKORTDDDLGWVLKHSNLSLEQLDLISLFGHESSVALALAF 326
D 235 SEKQSPDEYSGK-----IIVEMMRPGEETMPTAMTLAVK 270
OY 327 LQACRAVEELREHLEIARAKKELESELNWDYKKMDFTQCVINETLRLGNVRFHR 386
D 271 LSDNVALAKLVEENMEKRRRLTGE-EYKWTQVMSLFTQNVNITLRLMANIINGWR 329
OY 387 KALKDVRKGDIDPSGKVLVPLVISAHLNDRYDQNLFPNWRMOOQONNGASSSGSFS 446
D 330 KALKDVEIKGVLIPKGMVLASFISVHDEIDYDNPYOFDPRMDRINGSANSSIC--- 385
OY 447 TWGNMYMPEGGPRLCASSELAKEMAVFIHILVLFKFWELAEEDQPAFPVDPNGLP 506
D 386 -----FTPFGGGQRLCPLELSEKLETSIFLHLVLRYSM-TAEDEIVSFPVKKRRRLP 439
OY 507 IRVSR 512
D 440 IRVATV 445

RESULT 5
T48613
hypothetical protein F18022.190 - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cross)
C:Date: 20-Apr-2000 #sequence_revision 20-Apr-2000 #text_change 20-Apr-2000
C:Accession: T48613
R:Bevan, M.; Murphy, G.; Ridley, P.; Hudson, S.; Bancroft, I.; Mewes, H.W.; Rudd, S.; Le
submitted to the Protein Sequence Database, April 2000
A:Reference number: Z24493

```

```

A:Accession: T48613
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-382 <BEV>
A:Cross-references: EMBL:DB
A:Experimental source: cultivar Columbia; BAC clone F18022
C:Genetics:
A:Map position: 5
A:Introns: 31/2; 139/3; 204/1; 224/3; 253/3; 280/1; 315/3; 349/2
A:Note: F18022.190

Query Match 30.4%; Score 815; DB 2; Length 382;
Best Local Similarity 34.2%; Pred. No. 1,76-45;
Matches 159; Conservative 87; Mismatches 127; Indels 92; Gaps 4;

OY 45 GMPFLGETTGYLKPYTATTLGDFMOOHVSKYKITYSNLFGEPTIVSADAGINRETI 104
D 2 GMPFLGETTSPFKPHRSISIGFLDQVRSRGKVKFNSICGKAVYSCDQELNMTI 61
OY 105 GRLEPCSTPRISIGILGKWSMLVLDGMDHMDRSTISLNFISHARLTLTLDVERHTLV 164
D 62 GKLFPSDYPKAMHDLGKYSLLATGETHRKLNVIISFINLTKSPDPLHCAENLSISI 121
OY 165 LDSMOONSIFSAODEAKKFTFNLMARKHNSMDPGEETEOLAKREYTFPMKGVSAFLNP 224
D 122 LKSMNCEVEERHKEVKFTLSVMYNOLLSTKPEDPARLYVLDPLSTYWKGTSLPLP 181
OY 225 GTAYHAKALOSRATILKFERKMEERKLDIKEEDQEEBEVKTDEAEKSDHYRKORTD 284
D 182 GTGYTNAIKVSNRNRIHQNAIEDNNALREDFLDSIISND----- 224
OY 285 DLGKVLKHSNLSLEQLDLISLFGHESSVALALAFPLQACRAVEELREHLEI 344
D 225 -----EENAAI 230
OY 345 ARAKKESELEMDWDYKKMDFTQCVINETLRLGNVRFHRKALKDVRKGDIDPSGK 404
D 231 -RAKKGDEL-LNMDYOKMETQCVISEALRCGIYVTVHRAKHADHDKFNEYVIPKWK 288
OY 405 VLPVISAHLNDRYDQNLFPNWRMOOQONNGASSSGSFSFTWGNMYMPEGGPRLCAG 464
D 289 VEPITFAVHLDPSSLNENFEFPMRMTKT-----AFGGGVRCVCPG 329
OY 465 SELAKLEMAVFIHILVLFKFWELAEEDQPAFPVDPNGLDIRV 509
D 330 GELGKLIQIAFFLHLVLSYRWKIKSDEMPIAHPIYVEFRGMLEI 374

RESULT 6
T07859
cytochrome P450 homolog - tomato
C:Species: Lycopersicon esculentum (tomato)
C:Date: 14-May-1999 #sequence_revision 14-May-1999 #text_change 16-Feb-2001
C:Accession: T07859
R:Bishop, G.J.; Harrison, K.; Jones, J.D.
Plant Cell 8, 959-969, 1996
A:Title: The tomato Dwarf gene isolated by heterologous transposon tagging encodes th
A:Reference number: Z16181; MUID:96266705
A:Accession: T07859
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-464 <BS>
A:Cross-references: EMBL:U54770; NID:g1421740; PIDN:AA817070.1; PID:g1421741
A:Experimental source: strain GCR758
C:Genetics:
A:Gene: dwarf
C:Superfamily: Synechocystis cytochrome P450 slr0574; cytochrome P450 homology
F:273-436/Domain: cytochrome P450 homology <P45>

Query Match 28.4%; Score 761.5; DB 2; Length 464;
Best Local Similarity 33.4%; Pred. No. 6,26-42;

```

Matches 170; Conservative 90; Mismatches 188; Indels 61; Gaps 10;

12 LLLPSSLLPFLILKRRN--RKRFRNLPKSGMPFLGTTGKLPYATNTLG-DFM 68
 5 LIFLSSFGCLIFCALRRMNOVKNNKLPDGTGMPFLPFTETFLK-----LGPSFM 58
 QY 69 QOHVSKYKIRVSNLFGPPTIVSADAGLNFILONEGRFLFECSPRSIGLIGKMSMLV 128
 Db 59 KNORARYSPFRSHLIGCPITVSMDELNRYILVNEAGLVPGYQSMIDLKGNINAAV 118
 QY 129 VGDHMRDRSISLNFSLARLTLLKDYERHTLFLDSMOONSIFSQDEAKKFTENLM 188
 Db 119 NGSAKRYKRGALLSLISPTMIRDLPLPDEMRSHLTNM--DNKVYIDIQEKTNNKAFSS 177
 QY 189 AKHISMDPGEET---EQLKKEYTEPMKGVSAAPLNPGTAYHKAISOATILKFIERK 245
 Db 178 LKQI---AGLESTSLADEFMSEFNVLVGLTSLPILNPTNHYHGFQARKIYNLMLTL 233
 QY 246 MEERKLIDKEDEEVEEYKTEDEAEMSKSDHVRKORTDDLLGVLKHS---NLSTEOI 301
 Db 234 IEER-----ASKEIQHMDLGYLMNEATRFKLTDEM 266
 QY 302 LDLIISLFPAGHETSSVAIALAIFLQACPAVEELREHELEIARAKKEGSELNMDY 361
 Db 267 IDLIITILYSGEYETVSTSMAYKYLHDHPVLEELREKHAIREKKRP--EDPIDYNDY 324
 QY 362 KKMDETQCVINETLGNVVRFLHRKALKDVRKYDIPSGMKVLPVISAVALHDSRYDQ 421
 Db 325 RSMRTTRAVILETSKLATIVGVLKRTQDMEINCYIIPKGRITVYRELINYDRLYPD 384
 QY 422 PNLFPNRMQOONNGASSSGSFSFTMGNNYMPFGGPRLCAGSELAKLEMAVFIHNLVL 481
 Db 385 PYSFPMWMMDKS-----LEHONSFLVEGGGRQCPKRLGVAELSTFLHYVT 433
 QY 482 KFNWFLADDDQPAFPFPVDFPNGLPTRVS 510
 Db 434 KYRWEITGGDKLMKPRVPEANGLRIRVS 462

RESULT 7

cytochrome P450 - Arabidopsis thaliana
 N:Alternate names: protein T18B16.200; protein T5K18.10
 C:Species: Arabidopsis thaliana (mouse-ear cress)
 C:Date: 30-Apr-1999 #sequence_revision 30-Apr-1999 #text_change 28-Jul-2000
 A:Accession: T04444; T05806
 R:Bayan, M.; Benes, V.; Reichmann, S.; Borkova, D.; Ansoorge, W.; Bancroft, I.; Mewes, H.W.
 submitted to the Protein Sequence Database, April 1998
 A:Reference number: Z15359
 A:Accession: T04444
 A:Residues: 1-457 <BEV>
 A:Molecule type: DNA
 A:Cross-references: EMBL:AL021687
 R:Bayan, M.; Van Der Schueren, J.; Chuang, Y.J.; Veet, M.; Robben, J.; Voickaert, G.; Ba
 submitted to the Protein Sequence Database, April 1998
 A:Reference number: Z15453
 A:Accession: T05806
 A:Molecule type: DNA
 A:Residues: 131-457 <BEV>
 A:Cross-references: EMBL:AL022580
 A:Experimental source: cultivar Columbia; BAC clone T5K18
 C:Genetics:
 A:Map position: 4
 A:introns: 67/2; 173/3; 302/3; 358/1; 393/3
 A:Note: T18B16.200; T5K18.10
 C:Superfamily: Synchocystis cytochrome P450 slr0574; cytochrome P450 homology
 C:Keywords: chromoprotein; electron transfer; heme; iron; monooxygenase; oxidoreductase
 F:272-433/Domains: cytochrome P450 homology <P45>

Query Match 23.6%; Score 632; DB 2; Length 457;

Best Local Similarity 31.8%; Pred. No. 1.5e-33;
 Matches 151; Conservative 92; Mismatches 176; Indels 56; Gaps 12;

QY 13 LLLPSSLLPFLILKRRNKRFRNLPKSGMPFLGTTGKLPYATNTLGDFM 72
 Db 10 LPAQSLFLYFRCLISORFSSKLPDPTGMPYVET---FQLYSQDP--NVFFOSKO 65
 QY 73 SKYKIRVSNLFGPPTIVSADAGLNFILONEGRFLFECSPRSIGLIGKMSMLVLD 132
 Db 66 KRYGVSFTKHYLGCPCVWISSPEAKFVLYTKSHLFPKTPFASKERMLGKQAFPHQGDY 125
 QY 133 HRDMRSISLNFSLARLTLLKDYERHTLFLDSMOONSIFSQDEAKKFTENLM 192
 Db 126 HAKRLKLVRAFMEPSIRN--WVPIESIAQSLRSM--EGTMINITYQEKTYTFVALLSI 183
 QY 193 MSMDPGEET---EQLKKEYTEPMKGVSAAPLNPGTAYHKAISOATILKFIERK 249
 Db 184 F---GKDEVLYREDLKRCYIILKRGVSMFVNLPGLTFHSMKARRELSQLARILSER 239
 QY 250 KLIDKEDEEVEEYKTEDEAEMSKSDHVRKORTDDLLG--WVLKHSNLSTEOIDLILSL 308
 Db 240 R-----QNGSSH-----NDLGSFMGDKELTDQIADNITGV 272
 QY 309 LPAHETSSVAIALAIFLQACPAVEELREHELEIARAKKEGSELNMDYKKMDFQ 368
 Db 273 IFAARDTASVSMNLKYLAEINPNVLEAVTEQMAI--RKDEEGES--LTWGDTRKMLPTS 330
 QY 369 CVINETLGNVVRFLHRKALKDVRKYDIPSGMKVLPVISAVALHDSRYDQ 428
 Db 331 RVIOETLTVASLISFTFEAEVDEVEEGYLLPKGMKVLPLRLNIHSDIFSNGKFPDS 390
 QY 429 RMQOONNGASSSGSFSFTMGNNYMPFGGPRLCAGSELAKLEMAVFIHNLVL 483
 Db 391 RFE-----VAPKPRTEPMFGNGTHSCPGNELAKLEMSIMIHNLTTKY 432

RESULT 8

probable cytochrome P450 At2g29090 [Imported] - Arabidopsis thaliana
 N:Alternate names: cytochrome P450 homolog T914.17
 C:Species: Arabidopsis thaliana (mouse-ear cress)
 C:Date: 24-Mar-1999 #sequence_revision 24-Mar-1999 #text_change 16-Feb-2001
 A:Accession: T02739; D84692
 R:Rounsley, S.D.; Lin, X.; Kaul, S.; Shea, T.P.; Fujii, C.Y.; Mason, T.M.; Shen, M.;
 submitted to the EMBL data library, August 1998
 A:Description: Arabidopsis thaliana chromosome II BAC T914 genomic sequence.
 A:Reference number: Z14710
 A:Accession: T02739
 A:Status: translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-482 <ROU>
 A:Cross-references: EMBL:AC005315; NID:93461834; PID:93461849
 R:Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.
 M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; Vanaken, S.E.; Unayam, L.; Tallon,
 euss, D.; Niemman, W.C.; White, O.; Eilsen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter
 Nature 402, 761-768, 1999
 A:Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.
 A:Reference number: AB4420; MUID:20083487
 A:Accession: D84692
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-482 <STO>
 A:Cross-references: GB:AE002093; NID:93461849; PID:AMC33235.1; GSPDB:GN00139
 C:Genetics:
 A:Gene: At2g29090; T914.17
 A:Map position: 2
 A:introns: 80/2; 187/3; 238/3; 321/3; 351/3; 413/3; 453/2
 C:Superfamily: Synchocystis cytochrome P450 slr0574; cytochrome P450 homology
 F:291-453/Domains: cytochrome P450 homology <P45>

Query Match 22.8%; Score 611; DB 2; Length 482;

Best Local Similarity 29.8%; Pred. No. 3.5e-32;
Matches 153; Conservative 101; Mismatches 192; Indels 68; Gaps 13;

QY 16 PSLSLFLFLILK-----RRNRKTRNLPKSGMPFLGETIGYLPATTLGDMQ 69
D 20 PALLTLIVVVVLLKRWMLHMKOORLRLPGSMGLPIGET---RLTYENP-NSFFA 75
QY 70 QHVKYKIRYSNLFGEPTIVSADAGLNRFILONEGRLEFCSYPRSIGLIGKMSMLV 129
D 76 TRONKYGDIETHILGCVWISSPEARAVLSKAHLFPYPSKERMIGPEALFPHQ 135
QY 130 GDMRDM-RSISLNLFLSHARLTLKDVHRTFLVDSMOONSIFSADKAKFTENIM 188
D 136 GPVSTLKLRLVQSSFPMSALRPV--SHIELVLQTLSSWTSQKSINTLEVMKRYAFV 193
QY 189 AKHMSMDPGEE---TEOLKKEYVTMKVSAPLPGTAHYKALOSRATITIKFERK 245
D 194 ---TMSAFGDKDEEPTTIDVILKLYORLERGYSNPLDLPGLFHKSMKARIELSELRKV 250
QY 246 MEERKLDKEEDDEEVEKTEDEAEMSKSDHVRKQRTDDLLGWL-----KHSNLSTEQ 300
D 251 IEKRENGREE-----GGLGVLLAKKDKRNGLSQ 283
QY 301 ILDLILSLFAGHETSSVALAIFLQACPAAVEELREHELET-ARAKELGESELNMD 359
D 284 IADNLIGVFAATPTTASVLTWMLKYLHDHNLQEVSRQFSIRKTKKE--NRISME 341
QY 360 DYKMDTQVCYNETLRLGNVRLHKKALQVRYKGYDIPSGMKVLPVIAVHLDSRY 419
D 342 DTRKMPRLTRVIOETLRASVLSFTFRVADVEYDGLIKRGMKVLPLFRHHSSEF 401
QY 420 DQPLNFPMWMOQONNGSSGSGSFSWGNMYPEFGGPRLCAGSELAKTEMAVFIHHL 479
D 402 PDPEKFDSDRE-----VAKPYTYPFGNGVHSGSELAKTEMLLHL 448
QY 480 VLKFNWELAEDDQPFAP-PEVDFPNCGLPIRYSRI 512
D 449 TTSFRWEYIGDEGIQYCPFPVPRKKGLPIRVTP 482

RESULT 9
C71417
cytochrome P450 d13695c - Arabidopsis thaliana
N:Contains: oxidoreductase (EC 1.1.1.1)
C:Species: Arabidopsis thaliana (mouse-ear cress)
A:Variety: Columbia
C>Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 16-Jun-2000
C:Accession: C71417
R:Bevan, M.; Bancroft, I.; Bent, E.; Love, K.; Goodman, H.; Dean, C.; Berokamp, R.; Dirk
P.; Medler, H.; Medler, E.; Wambutt, R.; Weitzenegger, T.; Pohl, T.M.; Terry, N.; Giel
avanagh, T.; Hempel, S.; Kotter, P.; Entian, K.D.; Rieger, M.; Schaefer, M.; Funk, B.
Nature 391, 485-488, 1998
A:Authors: Mueller-Auer, S.; Silvey, M.; James, R.; Montfort, A.; Pons, A.; Pulgomech
erhoft, A.; Moore, T.; Jones, J.D.G.; Eneva, T.; Palme, K.; Benes, V.; Rechman, S.; An
C.; Chaitwalis, N.
A:Title: Analysis of 1.9 Mb of contiguous sequence from chromosome 4 of Arabidopsis thal
A:Reference number: A71400; MUID:86121113
A:Accession: C71417
A:Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-487 <BEV>
A:Cross-references: GB:297338; NID:g2244870; PIDN:CAB10309.1; PID:g2244888
C:Genetics:
A:Gene: d13695c
A:Map position: 4COP9-4G3845
C:Superfamily: Synchocystis cytochrome P450 slr0574; cytochrome P450 homology
C:Keywords: chromoprotein; heme; iron; metalloprotein; oxidoreductase
F:433/Binding site: heme iron (Cys) (axial ligand) #status predicted

Query Match 21.7%; Score 581; DB 1; Length 487;
Best Local Similarity 27.3%; Pred. No. 3.1e-30;
Matches 143; Conservative 100; Mismatches 205; Indels 76; Gaps 11;

QY 17 SLSLFLFLILKRRNRKTRNLPKSGMPFLGETIGYLPATTLGDMQHSYK 76
D 3 SLFLKTFHWWYQNRNPNTCKLPKSGMPFLGETIGTEFFEPHALDSTIKORVLEFF 62
QY 77 -----KIRSNLFGEPTIVSADAGLNRFILONEGRLEFCSYPRSIGI-----LG 121
D 63 ADFSINLSFRTSLFGKALISMDMLN-----LEMAKANSYPGYKVIIRFG 112
QY 122 KWSMLVLGDMHMRDRSISLNLFLSHARLTLKDVHRTFLVDSMOONSIFSADKAK 181
D 113 ENNFLOSKESSHKVRNLTFLQGLKSRMIDVDLARTMEGARGNYLDVKETSS 172
QY 182 KFTENLMAKHLIM-SMDPEEETEQKKEYTFPMGVSAPLNLTGTAHKL----- 232
D 173 KILGCLLKVWGEWEP--EAKELALCWRYFQSGWFERFNLPTGYKKMKVLPVOYT 230
QY 233 -----QSRATILKFERMEERKLDKEEDDEEVEVTEDEAEMSKSDHVRKQRTDDL 286
D 231 EADISWQARKKMKMLKRTVLTKRASGELEFFNIIIGEMGE----- 274
QY 287 LGWLKHSNLSTEQTLILSLFAGHETSSVALAIFLQACPAAVEELREHELETAR 346
D 275 -----GFTMSVENAVEYIYTFELVANETPRILAAVYKFTSDHKKVQKQLOREHEEIVR 328
QY 347 AKKELGESELNMDQYKMDPTQVCYNETLRLGNVRLHKKALQVRYKGYDIPSGMKV 406
D 329 GKAE-KEGGLTWEDKSNHFTQVINESLRITSTAPVLRLEHFOGDTTIPAGWTFM 387
QY 407 PVSAVHLDSNRDQPNLFNPMWMOQONNGSSGSGSFSWGNMYPEFGGPRLCAGSE 466
D 388 G-YPHIHNSKEDEPRYAFNPMWMEGKDLGALVS-----KTFIFGAGRILCVAE 437
QY 467 LAKEMAVFIHHLVLKFNWELAEDDQPFAPFPVDFPNCGLPIRYS 510
D 438 FAKQMAVFIHHL-FRYRMSKSGTTIRSPMLMPGCGDVOIS 480

RESULT 10
A84859
probable cytochrome P450 (imported) - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C>Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 23-Mar-2001
C:Accession: A84859
R:Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y
M.; Koo, H.; Mofat, K.S.; Cronin, L.A.; Shen, M.; Vanden, S.E.; Umayam, L.; Tallon,
euss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter
Nature 402, 761-768, 1999
A:Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.
A:Reference number: A84420; MUID:20083487
A:Accession: A84859
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-485 <STO>
A:Cross-references: GB:A8002093; NID:g4512670; PIDN:AAD21724.1; GSPDB:GN00139
C:Genetics:
A:Gene: At2g42850
A:Map position: 2
C:Superfamily: Synchocystis cytochrome P450 slr0574; cytochrome P450 homology
C:Keywords: heme; iron; metalloprotein
F:433/Binding site: heme iron (Cys) (axial ligand) #status predicted

Query Match 21.6%; Score 579.5; DB 2; Length 485;
Best Local Similarity 28.8%; Pred. No. 3.9e-30;
Matches 147; Conservative 107; Mismatches 203; Indels 53; Gaps 11;

QY 12 LLLPSLSLFL-----ILKRRNRKTRNLPKSGMPFLGETIGYLPATTL 64
D 14 LCIAATISSTLEFFRKKHNRRTKIKQKKR--LLPGEMGLPWIGETDIFYKAQKSNRVF 71
QY 65 GDFMOQHYSKYGKIRSNLFGEPTIVSADAGLNRFILONEGRLEFCSYPRSIGLIGKWS 124
D 65 GDFMOQHYSKYGKIRSNLFGEPTIVSADAGLNRFILONEGRLEFCSYPRSIGLIGKWS 124

Db 72 EDVNPRIKHNIFKTRIMSGPTIVNGAEANRLILSNEEFLVSSWPSSVOLMGNC 131
 QY 125 MLVLVGDHNRMRSLINFLSHARLTLLKDVENHFLFVDSMOQNSIFSAODEAKFT 184
 Db 132 IMAKGGERHVRGLVANSLSITGLSELTPLKCDVTKFHNHETEMGKEEISLYRSKALT 191
 QY 185 FNLMKAIH--MSMPGEDETEOLKKEYVTFMKGVASAPLPGTAHYHALQSRATILKEI 242
 Db 192 FTVEECLYGLKVEIGMEV-----FERVLEGVAFALVERPCSFARAKKARLEIEFLL 245
 QY 243 ERKMEERLIDKEDEQEEEVKTEDEAMSKSDHVRKORTDDLLGVNLKHSNSTEOIL 302
 Db 246 VGKREKREKREKGAE-----KPNITLFSRLVEELIKGV-----ITEEEVV 287
 QY 303 DLILSLFAGHETSSVALAIFLQACPKAVEELREELHETARAKKLGSE--LNMDDY 361
 Db 288 DNMVLVLAADHTTSYAMSMFKMLAQHPTCRDILLOHNAOI---KANGGCELTLYEDV 344
 QY 362 KKMDFQCVINETLRGLNVVRLHRRKALDVRKYGDIPSGMKVLPVISAHLNDSRYDQ 421
 Db 345 KMKYSMQVAVRETMRLSPRIFGSPKAVADIDYGYTIPKGMKILMTYGTNHYNPEIFQD 404
 QY 422 PNLNPRHMQOQNNGASSGSGSEFTGNCNNMPFGGGRCLAGSELAKLEMAVFIHLVL 481
 Db 405 PMSDFPTREKPIQAYT-----YLPFGGGRCLAGHOLAKISILVFMHEVYT 451
 QY 482 KFNNELEDOPFAFPVDFPN-GLPTRVS 510
 Db 452 GFDMSLVYPRDETISMDPLFPISLGMPIKIS 481

RESULT 11

H86185
 hypothetical protein [imported] - Arabidopsis thaliana
 C:Species: Arabidopsis thaliana (mouse-ear cress)
 C:Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 31-Mar-2001
 C:Accession: H86185
 R:Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso, J.; Chiu, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.; Hansen, N.F.; Hughes, B.; Hultz, L.
 Nature 408, 816-820, 2000
 A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.; C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maitl, R.; Marziani, Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
 A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, J.C.; Davis, R.W.; Ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
 A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
 A:Reference number: A86141; MUID:21016719
 A:Accession: H86185
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-490 <STO>
 A:Cross-references: GB:AE005172; NID:g238581; PIDN:AMB71462.1; GSPDB:GN00141
 C:Genetics:
 A:Map position: 1
 C:Superfamily: Synchocystis cytochrome P450 slr0574; cytochrome P450 homology

Query Match 21.6%; Score 579.5; DB 2; Length 490;
 Best Local Similarity 28.6%; Pred. No. 4e-30;
 Matches 133; Conservative 107; Mismatches 192; Indels 33; Gaps 9;
 QY 34 KTRFNLPGKSGMPFLGETTIGYLKPYATTLGDFMQOHVSKYGR--LYRSNLFGPEPTIVS 91
 Db 40 ENRHYLPDGLGWPFGICMLSLFAFKTSDDPSFTRLIKRYGPKGITYKAMFENPSITIV 99
 QY 92 ADAGLNRFILONEGRLEECSTPRISIGILGKWSMLVGVDMHRMRSISLNFSLHARLT 151
 Db 100 TTSOTCRVRLTDD--AFKPGWPTSMELIGRKSFGVISFEHKKRLRLTLARVNGHEALS 158
 QY 152 ILKLDVERHFLFVDSMOQNSIFSAODEAKFTFNLMKAIHMSMDPGEELTEOLKKEYVT 211
 Db 159 TYIPIEENVTIVLKDWTKMGEEFLLHRLKLTFRIL--MYTIFSSSEENVDALEREYTA 217

QY 212 FMKGVASAPLNLPGTAHYHALQSRATILKTERKMEERKLDIKEDEQEEEVKTEDEAM 271
 Db 218 LNYGRRAAVNIPGFAHYHARLAKRTLVAAFOSTYTER-----NOKKNIILSKKMDL 271
 QY 272 SKSDHVRKORTDDLLGVNLKHSNSTEOILDLILSLFAGHETSSVALAIFLQACP 331
 Db 272 DNLNVR-----DED-----GKTLDDPEIIDLVMYLNAGHESGHTIMATYFLOEHP 320
 QY 332 KAVEELREELHETARAKKELGESELMNDYKMKDFQCVINETLRGLNVVRLHRRKALD 391
 Db 321 EYLORAKAEQEMIIKSRPE--GQKGLSLKETRKMEFLSQVDETLRVITFSLTAFREACTD 379
 QY 392 VRYGYDIPSGMKVLPVISAHLNDSRYDQNLNPRHMQOQNNGASSGSGSEFTGNCN 451
 Db 380 VEMNGYILPCKMKVLTFRVDYHIDPEVPPDRKDDPARW---DNG-----FVPRAGA 428
 QY 452 YMPFGGGRCLAGSELAKLEMAVFIHLVLKFNNELEDOPFAF 496
 Db 429 FLPPGAGSHLCFPGNDLAKLEISIFLHFLTKYQVRSNPECPVNY 473

RESULT 12

T48973
 cytochrome P450-like protein - Arabidopsis thaliana
 N:Alternate names: protein Fl4D17.40
 C:Species: Arabidopsis thaliana (mouse-ear cress)
 C:Date: 02-Jun-2000 #sequence_revision 02-Jun-2000 #text_change 18-Aug-2000
 C:Accession: T48973
 R:Jordan, N.; Bangert, S.; Wiedemann, R.; Voss, H.; Unseld, M.; Mewes, H.W.; Rudd, S.
 submitted to the Protein Sequence Database, April 2000
 A:Reference number: 225008
 A:Accession: T48973
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-455 <JOR>
 A:Cross-references: EMBL:AJ353992; GSPDB:GN00061; ATSP:FL4D17.40
 A:Experimental source: cultivar Columbia; BAC clone Fl4D17
 C:Genetics:
 A:Gene: ATSP:FL4D17.40
 A:Map position: 3
 A:Insertions: 70/2; 178/3; 229/3; 312/3; 343/3; 346/1; 381/3
 C:Superfamily: Synchocystis cytochrome P450 slr0574; cytochrome P450 homology

Query Match 21.5%; Score 577; DB 2; Length 455;
 Best Local Similarity 28.9%; Pred. No. 5.2e-30;
 Matches 145; Conservative 84; Mismatches 173; Indels 100; Gaps 10;
 QY 32 NRKTRFNLPGKSGMPFLGETTIGYLKPYATTLGDFMQOHVSKYGR--LYRSNLFGPEPTIVS 91
 Db 28 NPKNSNGKLPPSGMGFPIIGETLDFPKPYGFEISPYLKKMLRVPLEFRTNIIAGVTVVS 87
 QY 92 ADAGLNRFILONEGRLEECSTPRISIGILGKWSMLVGVDMHRMRSISLNFSLHARLT 151
 Db 88 TDKDVNMELIRQENKSFSLSPDGLMKPRKDSFLKGINHKKIKQITLLHLLSEGLK 147
 QY 152 ILKLDVERHFLFVDSMOQNSIFSAODEAKFTFNLMKAIHMSMDPGEELTEOLKKEYVT 211
 Db 148 KILKMDVTRREHLSSAKTGRLDVDAVSKLITAHLLPKMMS-----NLKPGTQA 198
 QY 212 FMKGV-----VSAPLNLPGTAHYHALQSRATILK--FIERKMEERKLDIKE 255
 Db 199 KLMDIFAFPFDMWRRTSYLSAGKGLNTLW--ACRGMREIKIYIMRKTSSEKY----- 252
 QY 256 EDOEEEVKTEDEAMSKSDHVRKORTDDLLGVNLKHSN-----LSTEOILDLILSLF 310
 Db 253 -----DDFLNFAIESEKAGELLNENALITLIFLISC 284
 QY 311 AGHTSSVALAIFLQACPKAVEELREELHETARAKKELGESELMNDYK--KMDFTQC 369
 Db 285 VTQDTSKAICLAVKFLLENPKVLAELKKEH--EVLESREDEKGVWEEYRHKMFTFN- 342

QY 370 VINETLRLGNVYRFLHRRALADVRKGYDIPSGMKVLPVISAHLNDSRYDPLNPNR 429
 Db 343 -----MKSGYTPAGWIMTIPSVVHDPELTENPFEENPNR 379
 QY 430 WQOQNGASSSGSFSWGNMMPFGGPRLCAGSELAKEMAVFIHHLVTKFMWELAE 489
 Db 380 WEGKELRAGS-----KTFMWFGTGLRQCAAGAEFAQDISVFLHLLVTTYNFSLHQ 429
 QY 490 DDQPAFPFVDFPNCPLPIRVS 511
 Db 430 DCEVLRAVRAHLNPGNISINISK 451

RESULT 13

884733
 Probable cytochrome P450 (imported) - Arabidopsis thaliana
 C:Species: Arabidopsis thaliana (mouse-ear cress)
 C:Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 16-Feb-2001
 C:Accession: B84733
 R:Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.;
 M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; Vanhaken, S.E.; Umayam, L.; Tallon,
 L.; Nierman, W.C.; White, O.; Elsen, J.A.; Salzberg, S.L.; Fraser, C.M.; Ventner,
 Nature 402, 761-768, 1999
 A:Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.
 A:Reference number: A84420; MUID:20083487
 A:Accession: B84733
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-489 <STO>
 A:Cross-references: GB:AE002093; NID:q3831452; PIDN:AAC69934.1; GSPDB:GNO0139
 C:Genetics:
 A:Gene: At2g32440
 A:Map position: 2
 C:Superfamily: Synecocystis cytochrome P450 slr0574; cytochrome P450 homology

Query Match 20.4%; Score 547.5; DB 2; Length 489;
 Best Local Similarity 27.6%; Pred. No. 4,6e-28;
 Matches 146; Conservative 105; Mismatches 211; Indels 67; Gaps 14;

QY 8 TLPLPLLSLSLLFLI--LKRNR-----RKTRENLPGKSGMPFLETGYLK 57
 Db 4 TGLIMMPF-LIILGLEVLKWLKRVNWIYKSGKEKKHYLPBGDLGWPVIGNMMSFLR 62
 QY 58 PYTATLDFMQOHKSKG--IYNSNLFGEPTIYSADAGLRFILQNGRLEFESYPS 115
 Db 63 AKTSDPESEFISYTRRGRTGIRAHMGVPCVLTTPETCRVLTDD-AFHIGMPKS 121
 QY 116 IGGILGKMSMLVAGDMHDMRSISNLFSHARLRTILKDYERHTLFVLDSDMOQSIFS 175
 Db 122 TMKILGRKSFVGSFPEHRLRLRLTSAPVNGPEALSVIQFTEBYVNDLEKMSKGELE 181
 QY 176 AODEAKKFTFNMAKHMSMDPEEETEQLKREYTFEMKGVASAPLNLPGTAYHKAQSR 235
 Db 182 FLSHRKLFLFYI-MYIFLSESEHVMDSLEREYTNLNVGRAMGINLPGFAYHRAKAR 240
 QY 236 ----ATILKTRKMEERKLDI--KEEOEVEEVKTEDEAKMSKSHVAKQRTDDLLGM 289
 Db 241 KKLVAFOSIVNRRNRKONISSNRKMDLNLIDVKE-----NGRVLD--- 286
 QY 290 VLKHSNLSLEQILDLILSLFAGHETSSVALAIFFLQACRAVEELREHELEIARAK 349
 Db 287 -----EELIDILLMLNNGHSSGHLTMTATILMQHEHMILOKAKEEBERT--VK 335
 QY 350 ELGSELNMDYKKMDFTCVINEITLRLGNVYRFLHRRALKDVRKGYDIPSGMKVLPV 409
 Db 336 RAPGOKLTAKTERENVYLSQVIDETLRTVTFSLTAFREAKSDVQMDGYIIPGKMKVLTWF 395
 QY 410 SAVHLDNSRYDPLNPNRMOQOONNGASSSGSFSWGNMMPFGGPRLCAGSELAK 469
 Db 336 RNWHLDPPEIYPPKPKFDDSRWE-----GTPRAGTFLLPGGLSHLCGNDLAK 443
 QY 470 LEMAVFIHHLVLPFMWELAEDDQPAFPFVDFPNCPLP-----IRVSRL 513

Db 444 LEISIFLHHFLKRYVERSNPGCPVAF-----LPHNRKONCLARTRTM 488

RESULT 14

D96813
 Hypothetical protein T30F21.17 (imported) - Arabidopsis thaliana
 C:Species: Arabidopsis thaliana (mouse-ear cress)
 C:Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 31-Mar-2001
 C:Accession: D96813
 R:Rheologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alon,
 Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar,
 ansen, N.F.; Hughes, B.; Hulzar, L.
 Nature 408, 816-820, 2000
 A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim,
 C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luoro, J.S.; Maiti, R.; Matzla,
 Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
 A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallo,
 ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Ventner, J.C.; Davis, R.W.
 A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
 A:Reference number: A86141; MUID:21016719
 A:Accession: D96813
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-460 <STO>
 A:Cross-references: GB:AE005173; NID:q4836883; PIDN:AAD30586.1; GSPDB:GNO0141
 C:Genetics:
 A:Gene: T30F21.17
 A:Map position: 1

Query Match 18.6%; Score 497.5; DB 2; Length 460;
 Best Local Similarity 26.2%; Pred. No. 7,3e-25;
 Matches 133; Conservative 102; Mismatches 205; Indels 67; Gaps 12;

QY 13 LPLPSLSLSLLFLI--LKRNR-----RKTRENLPGKSGMPFLETGYLKPYTATLDFMQOHV 72
 Db 9 MLMVALVYRISHWILRWNSNPKPCPKLPFGSGMGPFLIGETLDFEKKCGVEGPTPKKRM 68
 QY 73 SKYGIKYSNLFGEPTIYSADAGLNFILQNGRLEFESYPSISGIGLAKMSMLVAGDM 132
 Db 69 IRGGLFRTNIGSKTVASTDPDVIIHQIFROENTSPELGYPDIFVAVFGKDNILFKEVFI 128
 QY 133 HRDMRSISNLFSHARLRTILKDYERHTLFVLDSDMOQSIFSADQDEAKKFTFNMAKH 192
 Db 129 HRYLOKITMOLIGSEGLKQTMGLGMDKATRDHIRSIASOGSFVNRKVENLVAVAYTPRL 188
 QY 193 MSMDPEEETEQLKREYTFEMKGVASAPLNLPGTAYHKAQSRATILKTRKMEERKLD 252
 Db 189 ISNKRPEQSKLIDN-----LNAFNLDMPKSF-LRLSTWKAVTRAKSR----- 231
 QY 253 IKEEDQEEVEVTEDEAKMSKSHVAKQRTDDLLGWLKH-----SNLSTQIIDLILS 307
 Db 232 -----EQAIVQMKVLMRRKRETKQ---EDFLNTLLELEKDSFPDQSAIHLPL 281
 QY 308 LFLAGHETSSVALAIFFLQACRAVEELREHELEIAAKKELGSELNMDYK-KMPF 366
 Db 282 LAFALREGTSCCTALAVKFIKDPKVLAKREHKAIVNRRD-KEAGVSWEERYHNHMF 340
 QY 367 TOCVINEITLRLGNVYRFLHRRALKDVRKGYDIPSGMKVLPVISAHLNDSRYDPLN 425
 Db 341 TMMVSNEVRLANTPPLRLKRAVDVEIKGY-----YLEVTHMGFG----- 382
 QY 426 NPMRMOQOONNGASSSGSFSWGN-NYMPFGGPRLCAGSELAKEMAVFIHHLVTKFN 484
 Db 383 --WLMQ-----GKEMTWGSKTFMAFGVGLCVGAESFRLOMALFIHHLVAYVD 429
 QY 485 WELAEEDQPAFPFVDFPNCPLPIRVS 511
 Db 430 FSMVDSEITIRSPHOYTQDLINISQ 456

RESULT 15

T02263
cytochrome P450 DWARF3 - maize
N:Contains: oxidoreductase (EC 1.-.-.-)
C:Species: Zea mays (maize)
C:Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 21-Jul-2000
C:Accession: T02263
R:Winkler, R.G.; Helentjaris, T.
Plant Cell 7, 1307-1317, 1995
A:Title: The maize dwarf3 gene encodes a cytochrome P450-mediated early step in gibberellin biosynthesis
A:Reference number: Z14648; MUID:96004534
A:Accession: T02263
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Residue type: mRNA
A:Residues: 1-519 <MIN>
A:Cross-references: EMBL:U32579; NTD:9987266; PIDN:AAC49067.1; PID:9987267
A:Experimental source: strain B73
C:Genetics:
A:Gene: dwarf3
C:Function:
A:Description: Involved in an early step in gibberellin biosynthesis
A:Pathway: gibberellin biosynthesis
C:Superfamily: Synchocystis cytochrome P450 slr0574; cytochrome P450 homology
C:Keywords: oxidoreductase
I:325-468/Domain: cytochrome P450 homology <P45>

| | | | | |
|-----------------------|------------------|--------------------|------------|-------------|
| Query Match | 18.38; | Score 491.5; | DB 1; | Length 519; |
| Best Local Similarity | 28.28; | Pred. No. 2.1e-24; | | |
| Matches 129; | Conservative 91; | Mismatches 189; | Indels 49; | Gaps 11; |

[illegible]

Search completed: October 17, 2001, 22:08:02
Job time: 8206 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 17, 2001, 22:08:06 ; Search time 30.88 Seconds

(Without alignments)
569.076 Million cell updates/sec

Title: US-09-502-426-2
Perfect score: 2681
Sequence: 1 MFETEHNTLLPLLLPLSLLS.....FAFPVDPNGLPIRVSRLL 513

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 93435 seqs, 34255486 residues
Total number of hits satisfying chosen parameters: 93435

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_39:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description |
|------------|-------|-------------|--------|----|--------------------|
| 1 | 978.5 | 36.5 | 472 | 1 | C901_ARATH |
| 2 | 761.5 | 28.4 | 464 | 1 | CP85_LYCES |
| 3 | 579.5 | 21.6 | 490 | 1 | C883_ARATH |
| 4 | 503 | 18.8 | 492 | 1 | CP26_BRARE |
| 5 | 491.5 | 18.3 | 519 | 1 | C881_MAIZE |
| 6 | 461 | 17.2 | 497 | 1 | CP26_MOUSE |
| 7 | 438 | 16.3 | 497 | 1 | CP26_HUMAN |
| 8 | 324.5 | 12.1 | 520 | 1 | YRV2_CAEEL |
| 9 | 321.5 | 12.0 | 518 | 1 | YRVA_CAEEL |
| 10 | 316 | 11.8 | 500 | 1 | C912_ARATH |
| 11 | 312.5 | 11.7 | 504 | 1 | CP32_RAT |
| 12 | 312 | 11.6 | 501 | 1 | CPJ5_MOUSE |
| 13 | 310.5 | 11.6 | 520 | 1 | YRV1_CAEEL |
| 14 | 303 | 11.3 | 501 | 1 | CPJ6_MOUSE |
| 15 | 302 | 11.3 | 494 | 1 | CPA8_MESAU |
| 16 | 299.5 | 11.2 | 492 | 1 | CPBC_RAT |
| 17 | 297 | 11.1 | 491 | 1 | CPB9_MOUSE |
| 18 | 294.5 | 11.0 | 491 | 1 | CPB4_RABIT |
| 19 | 293 | 10.9 | 491 | 1 | CPB1_RAT |
| 20 | 291.5 | 10.9 | 530 | 1 | CP51_YEAST |
| 21 | 290.5 | 10.8 | 502 | 1 | CP83_ARATH |
| 22 | 289 | 10.8 | 503 | 1 | CP39_RAT |
| 23 | 288.5 | 10.8 | 492 | 1 | CPB1_MOUSE |
| 24 | 288.5 | 10.8 | 503 | 1 | CP33_HUMAN |
| 25 | 288.5 | 10.8 | 538 | 1 | CP18_DROME |
| 26 | 288 | 10.7 | 459 | 1 | CPXN_ANASP |
| 27 | 287.5 | 10.7 | 520 | 1 | YRV5_CAEEL |
| 28 | 287 | 10.7 | 506 | 1 | CPJ2_HUMAN |
| 29 | 286.5 | 10.7 | 502 | 1 | CPJ1_CANCA |
| 30 | 286.5 | 10.7 | 533 | 1 | C762_SOLME |
| 31 | 284.5 | 10.6 | 505 | 1 | CP3G_MOUSE |
| 32 | 281.5 | 10.5 | 504 | 1 | CP51_CANTR |
| 33 | 281 | 10.5 | 528 | 1 | P14263 candida tro |

| | | | | | | |
|----|-------|------|-----|---|------------|---------------------|
| 34 | 280.5 | 10.5 | 504 | 1 | CP3B_MOUSE | O64459 mus musculus |
| 35 | 280 | 10.4 | 499 | 1 | C771_SOLME | P37123 solanum mel |
| 36 | 280 | 10.4 | 500 | 1 | CPJ1_RABIT | P52786 oryctolagus |
| 37 | 279.5 | 10.4 | 513 | 1 | C773_SOYBN | O48928 glycine max |
| 38 | 279.5 | 10.4 | 519 | 1 | YRV8_CAEEL | O27520 caenorhabdi |
| 39 | 279 | 10.4 | 501 | 1 | CAD2_DROME | O27589 drosophila |
| 40 | 279 | 10.4 | 502 | 1 | CPJ3_RAT | P51590 rattus norv |
| 41 | 278.5 | 10.4 | 501 | 1 | CPJ6_RABIT | P11707 oryctolagus |
| 42 | 278 | 10.4 | 470 | 1 | CPBK_MOUSE | O62397 mus musculus |
| 43 | 278 | 10.4 | 503 | 1 | CP3D_MOUSE | O64464 mus musculus |
| 44 | 277.5 | 10.4 | 491 | 1 | CPB5_RABIT | P12789 oryctolagus |
| 45 | 276 | 10.3 | 491 | 1 | CPB2_RAT | P04167 rattus norv |

ALIGNMENTS

| RESULT | ID | STANDARD | PRT | 472 AA. |
|--|--|----------|-----|---------|
| 1 | C901_ARATH | | | |
| AC | Q42569 | | | |
| BT | 15-DEC-1998 (Rel. 37, Created) | | | |
| DT | 15-DEC-1998 (Rel. 37, Last sequence update) | | | |
| DT | 15-DEC-1998 (Rel. 37, Last annotation update) | | | |
| DE | CYTCHROME P450 90A1 (EC 1.14.-.-) | | | |
| GN | CYP90A1 OR CYP90 OR CPD. | | | |
| OS | Arabidopsis thaliana (Mouse-ear cress). | | | |
| OC | Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta; | | | |
| CC | Magnoliophyta; Eudicotyledons; core eudicots; Rosidae; eurosids II; | | | |
| CC | Brassicales; Brassicaceae; Arabidopsis. | | | |
| OX | NCBI_TaxID=3702; | | | |
| RN | [1] | | | |
| RP | SEQUENCE FROM N.A. | | | |
| RC | STRAIN=CV. COLUMBIA. | | | |
| FX | MEDLINE=96200769; PubMed=8612270; | | | |
| RA | Szekeeres M., Nemeth K., Koncz-Kaiman Z., Mathur J., Kauschmann A., | | | |
| RA | Altman R., Redei G.P., Nagy F., Schell J., Koncz C.; | | | |
| RT | "Brassinosteroids rescue the deficiency of CYP90, a cytochrome P450, | | | |
| RT | controlling cell elongation and de-etiolation in Arabidopsis."; | | | |
| RL | Cell 85:171-182(1996). | | | |
| CC | -I- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY. | | | |
| CC | ----- | | | |
| CC | This SWISS-PROT entry is copyright. It is produced through a collaboration | | | |
| CC | between the Swiss Institute of Bioinformatics and the EMBL outstation - | | | |
| CC | the European Bioinformatics Institute. There are no restrictions on its | | | |
| CC | use by non-profit institutions as long as its content is in no way | | | |
| CC | modified and this statement is not removed. Usage by and for commercial | | | |
| CC | entities requires a license agreement (See http://www.isb-sib.ch/announce/ | | | |
| CC | or send an email to license@sib-sib.ch). | | | |
| CC | ----- | | | |
| DR | EMBL; X87367; CA60793.1; - | | | |
| DR | EMBL; X87368; CA60794.1; - | | | |
| DR | InterPro; IPR001128; - | | | |
| DR | Pfam; PF00067; P450; 2. | | | |
| DR | PROSITE; PS00086; CYTOCHROME_P450; 1. | | | |
| KW | Oxidoreductase; Monooxygenase; Heme. | | | |
| FT | BINDING 418 418 HEME (BY SIMILARITY) | | | |
| SO | SEQUENCE 472 AA; 53785 MW; 41A73F46D64E34F CRC64; | | | |
| Query Match | 36.5%; Score 978.5; DB 1; Length 472; | | | |
| Best Local Similarity | 41.9%; Pred. No. 1.3e-55; | | | |
| Matches 210; Conservative 79; Mismatches 171; Indels 41; Gaps 8; | | | | |
| OY | 12 LLLPLSLLLFLLKRRNRKTRFNLPPCKSGMPFLGFTGYLKPYATATLGGFMQOH 71 | | | |
| DB | 7 LLLSSIAAGFL--LLRRTYRRMGGLPGSLGLIGFTFOLLGAVKTEMPDFIDER 63 | | | |
| OY | 72 VSKYKTRNSNFCGPTVVSADAGNRIQLNEGSLFECSPRSIGGLKSKMVLVGD 131 | | | |
| DB | 64 VARYGVFTFHLFGPTTFVSADPETNRFVLONEGLFECSPASICNLGKHSLLLMGMS 123 | | | |
| OY | 132 MHRDRSISLNLFLSHARLRTLLKDKVERHTLFLVDSMOONSIFSAODBAKKFTFNLMAKH 191 | | | |

```

Db 124 LHKRHSITMSFANSIIKIDMLMDIDLVFNIDSMSSVLL--MEBAKKTFFELFYKQ 181
Qy 192 IMSMDPGEETEOAKKEVTPMKGVASAPLNLPGTAYKALQSRATILKFERKMEERKL 251
Db 182 LMSDPPG-EMWESLKRKEVLVIEGFSLPLPLFSTYRKALQAR-----RKVAELT 232
Qy 252 DIKEEDOEFEVEKTEDEAEMSKSDHVRKQRTDDLLGVNLHNSLSTEQILDLISLFLA 311
Db 233 VVWKKRREBEAGAE-----RKDMIAALLAADGDSDEIDYLVALLVA 278
Qy 312 GHETSSVALAIAIFLQACPAVEELREHELEIARAKKEGSESELNMDYKMDFTQCVI 371
Db 279 GYETSTTMTLAVKFLFETPLALQKHEHEKIRAMKSD--SYSEMEDYKSMPEFTQCVV 336
Qy 372 NETLRGLVAVFLRKALKDVRKGYDIPSGMKVLPVSAVHLNDSRYDQNLNPNRMQ 431
Db 337 NETLRVAVIIGGVRRATDVEIKGYKIPKGMKVSSFRVAHDLPNHKKDARTNPNRMQ 396
Qy 432 QONNGASSSGSFSFTWGN--YMPFGGPRLCAGSELAKLEMAVEIHLVLKFMNELAED 490
Db 397 -----SNSVTTGGSNVFTFPGGPRLCPGVELARVALSVFLHRLVYGFSSVPAEQ 446
Qy 491 DQPAFPEVDFPENGPIRVSR 511
Db 447 DKLVFEPTTRQKRYPIFKR 467

RESULT 2
CP85_LYCES STANDARD: PRT: 464 AA.
ID CP85_LYCES 043147;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DE CYTOCHROME P450 85 (EC 1.14.-.-) (DWARF PROTEIN).
GN CP85 OR D.
OS Lycopersicon esculentum (Tomato).
OC Eukaryota: Viridiplantae: Embryophyta: Tracheophyta: Spermatophyta:
OC Magnoliophyta: eudicotyledons: core eudicots: Asteridae I:
OC Solanales: Solanales: Solanum.
OX NCBI_TaxID=4081;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. GCR758;
RX MEDLINE=96266705; PubMed=6672892;
RA Bishop G.J., Harrison K., Jones J.U.G.D.;
RT "The tomato Dwarf gene isolated by heterologous transposon tagging
encodes the first member of a new cytochrome P450 family.";
RL Plant Cell 8:959-969(1996).
CC -1- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: U54770; AAB17070.1; -
CC InterPro: IPR001128; -
CC Pfam: PF00067; P450.1.
CC PRINTS: PR00385; P450.
CC PROSITE: PS00086; CYTOCHROME_P450.1.
CC Oxidoreductase; Monooxygenase; Membrane; Heme.
CC BINDING 414 414 HEME (BY SIMILARITY).
CC SEQUENCE 464 AA; 53706 MW; D2B21AAB7B14E94 CRC64;

```

```

Qy 12 LLLPLSLLLFLILLKRRN--RKTRPNLPBGKGMPELETIGLYKPYATTLG-DFM 68
Db 5 LIFLSFFGLCICFTALLRNQVYKNOKNLPRTGMNPLREGETTEFLK-----LGSEFM 58
Qy 69 QOHVSKYKTIYRSLNFEPTIVSADAGINREPLTONEGRLEFCSPRSIGLIGKWSMLV 128
Db 59 KNQARVGFSPFKSHLCPTPLSDSELNRYILVNEAKGLVPGVQSMIDILGKCNIAAV 118
Qy 129 VGDHNRDQRSISLFLSHARKITLLDKDVERHTLFVLDNQONSIFSAQDARKFTFLM 188
Db 119 NSAKKTYMKGALLSLISPTMRDQLLPKIDFPMKSHLTNW-DNKVIDIQETNNMAFLSS 177
Qy 189 AKHIMSMDPGEET---EOLKKEYVTFMKGVASAPLNLPGTAYKALQSRATILKFERK 245
Db 178 LKQI-----AGIESITSLAQEFMSEFFNLVGLSLPILNPNRYNRGQARKIYNLRTL 233
Qy 246 MEERKLDIKEEDOEFEVEKTEDEAEMSKSDHVRKQRTDDLLGVNLHNS---NLSTEQI 301
Db 234 IEERR-----ASKEIQHDMLYLIMEEATRFKLTDEM 266
Qy 302 IDLLSLFLPAGHETSSVALAIAIFLQACPAVEELREHELEIARAKKEGSESELNMDY 361
Db 267 IDLIITLISGYETVSTSMNAVYLLHDHPKVLDELREHMAIEKKRP--EDPIDNDY 324
Qy 362 KKMFTQCVINETLRGLNVAVFLRKALKDVRKGYDIPSGMKVLPVSAVHLNDSRYDQ 421
Db 325 RSMRTFRAVILETSRLATIVNGVLRLKTTQDMINGIYIIPKMKRIYVTRLYNTPRLYPD 384
Qy 422 PNLFPNRMQOONNGASSSGSFSFTWGN--YMPFGGPRLCAGSELAKLEMAVEIHLVL 481
Db 385 PYSEFNPWRMDKS-----LEHONSFLVFGGGRQCPGKELGVAIESTFLHAFVT 433
Qy 482 KFMNELAEDDQPAFPEVDFPENGPIRVSR 510
Db 434 KYRWEIIGDKLMEPPEVAPNGLIRVS 462

RESULT 3
C883_ARATH STANDARD: PRT: 490 AA.
ID C883_ARATH 023051;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DE CYTOCHROME P450 88A3 (EC 1.14.-.-).
GN CYP88A3 OR YUP8H12.23.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota: Viridiplantae: Embryophyta: Tracheophyta: Spermatophyta:
OC Magnoliophyta: eudicotyledons: core eudicots: Rosidae: eurosids II;
OC Brassicales: Brassicaceae: Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. COLUMBIA;
RA Theologis A., Osborne B.L., Vysotskaia V.S., Federspiel N.A.,
RA Tortumli M., Yu G., Oji O., Araujo R., Chung R.W., Dewar K., Dietrich F.,
RA Ecker J.R., Marzilli A., Oefner P., Davis R.W.;
RL Submitted (SEP-1997) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: AC000098; AAB71462.1; -
CC InterPro: IPR001128; -
CC Pfam: PF00067; P450.1.
CC PRINTS: PR00385; P450.

```

```

Query Match 28.4%; Score 761.5; DB 1; Length 464;
Best Local Similarity 33.4%; Pred. No. 9.5e-42;
Matches 170; Conservative 90; Mismatches 188; Indels 61; Caps 10;

```

DR PROSITE; PS00086; CYTOCHROME P450; 1.
KW Oxidoreductase; Monooxygenase; Transmembrane; Heme.
FT TRANSMEM 6 POTENTIAL.
FT BINDING 439 439 HEME (BY SIMILARITY).
SQ SEQUENCE 490 AA; 56409 MW; 7FD8CD7A8864D298 CRC64;

Query Match 21.6%; Score 579.5; DB 1; Length 490;
Best Local Similarity 28.6%; Pred. No. 4.3e-30;
Matches 133; Conservative 107; Mismatches 192; Indels 33; Gaps 9;

QY 34 KTRFNLPPGKSGWPLFGTIGYKATTTGLDFMQOHVSKYK--IYRSNLFGEPTIVS 91
DB 40 ENRHVLPDGLGMPITGMNLSFLRAFKTSDDPSFTRLIKRGPRGIVKAIHNGPSTIV 99
QY 92 ADAGINRFTLDNEGRLFECSYPRSTIGLIGKSMVLVGDHMRMRSISLNFSLARLFT 151
DB 100 TTSDTCRRVLTDD--AFKPGWPTSMELIGKRSFVGISFEHKRLRLTLAAVNGHEALS 158
QY 152 ILTKVREHTLFLVDSMOONSIFSAODEAKKPTFNLMKHMSPGCEETQLKKEYVT 211
DB 159 TYIPTEENVITVLDKWKTKMGEEFTLHRLKLFRTI--MYTFLSESEVMDALEREYTA 217
QY 212 FMKGVVASPLNPGTAYHKALOSRAITILKTERKMEERKLDIKEEDOEVEEYKTEDEAM 271
DB 218 LNYGVRAVAVNIPGFAYHRAKAKTIVAARQSYTER--NORONLTSKKDKL 271
QY 272 SKSDHYRQRDLDLGLVHKSINLSTEOILDILSLFAGHETSSVAIALAIFLQACP 331
DB 272 DNLNVK-----DGD-----GKTLDEEIIIDVLLMYNAGHSSGHTIMATVFLQEH 320
QY 332 KAVELREHLEIARAKKELESELNMDYKMDTQCVINETLRIGNVRLHRAKLD 391
DB 321 EYLQAKAKDEQELIKSRP--GQKGLSKETRKMEFLSOVYDDELTVTFSLAFREAKTD 379
QY 392 VRYKGYDIPSGMKVLPVISAHLNDSRYDQPNLFNPMRMOQONGASSSGSFGTWN 451
DB 380 VEMNGYLIPKGMKVLTPWRDVHIDPEVFPDPKFPARM--DNG-----FVPKGA 428
QY 452 YMEFGGPRLCAGSELATLEMAVFIHHLVLFKNMELAEDDQFAP 496
DB 429 FLPEAGSHLCPGNDLAKLEISIFLHFLKLYQVRSNDECPVMY 473
RESULT 4
CP26_BRAE STANDARD: PRT: 492 AA.
AC P79739;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE CYTOCHROME P450 26 (EC 1.14.--) (RETINOIC ACID-METABOLIZING
CYTOCHROME) (P450RA1) (RETINOIC ACID 4-HYDROXYLASE).
GN CYP26.
OS Brachydanio rerio (Zebrafish) (Zebra danio).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Cyprinidae; Teleostei; Euteleostei; Ostariophysi;
OC Cypriniformes; Cyprinidae; Rasbora; Danio.
OX NCBI_TaxID=7955;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=97094702; PubMed=8939936;
RA White J.A., Guo Y.-D., Baetz K., Beckelt-Jones B., Bonasoro J.,
RA Hsu K.E., Dilworth F.J., Jones G., Petkovitch M.;
RT "Identification of the retinoic acid-inducible all-trans-retinoic
RT acid 4-hydroxylase."
RL J. Biol. Chem. 271:2922-2927(1996).
CC -!- FUNCTION: PLAYS A KEY ROLE IN RETINOIC ACID METABOLISM. ACTS ON
CC RETINIDS, INCLUDING ALL-TRANS-RETINOIC ACID (RA) AND ITS
CC STEREOISOMER 9-CIS-RA. CAPABLE OF 4-HYDROXYLATION, HYDROXYLATION,
CC RESPONSIBLE FOR GENERATION OF SEVERAL HYDROXYLATED FORMS OF RA,
CC INCLUDING 4-OH-RA AND 4-OXO-RA.
CC -!- SUBCELLULAR LOCATION: MEMBRANE-BOUND. ENDOPLASMIC RETICULUM.

CC -!- INDUCTION: BY RETINOIC ACIDS (RA).
CC -!- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL; U68234; AAC60045.1; -;
DR ZFIN; ZDB-GENE-990415-44; cyp26.
DR InterPro; IPR001128; -;
DR Pfam; PF00067; P450.1;
DR PRINTS; PR00385; P450.1.
DR PROSITE; PS00086; CYTOCHROME_P450; 1.
KW Oxidoreductase; Monooxygenase; Membrane; Heme; Microsome;
KW Endoplasmic reticulum.
FT BINDING 438 438 HEME (BY SIMILARITY).
SQ SEQUENCE 492 AA; 56281 MW; FD471435B2F30509 CRC64;

Query Match 18.8%; Score 503; DB 1; Length 492;
Best Local Similarity 27.6%; Pred. No. 3.4e-25;
Matches 144; Conservative 95; Mismatches 216; Indels 66; Gaps 14;

QY 7 HTLLPLLLPSLLSLFL-----ILKRRNRKTRNLPFGSGWPLFGTIGYKLP 58
DB 4 YLLMTFLGTLVPLVFLAIAVKLMELMIRRYDPCRSPLPGTGLFGTIGTLII-- 61
QY 59 YATITLGDVMOOHVSKGYGIVNSLFGPTIYSADAGLNRLFIQNGRLFECSYPRSTIG 118
DB 62 ---LORRKFLRMKROQYGGCYITHTLGNPTVYMGADVNOIILGHRKLVSYQWPAVST 118
QY 119 ILGKSMVLVQDMRD-----MRSISLNFSLHARLRTILKDVREHTLFLVDSW--QONS 172
DB 119 ILGSDTLNVHGVQKRNKKKAIMRAFSRALDH-----YIVIQGVASAIQEWLQKXS 172
QY 173 TFSADAKKFTFNLMKHMSPGCEET--EQLKKEYVTNKGVSAPLPGTAYHK 230
DB 173 CVLVYPEMKKLMFRIAMRLIFGEFEQIKTDQELVAFEEIMKNIPLIDVPFSGLYR 232
QY 231 ALQSRATILKTERKMEERKLDIKEEDOEVEEYKTEDEAMKSDHYRQRDLDLGLV 290
DB 233 GLRAR---NFIHSKLEENIRKKIODDQNEQKXKDALQL---IENSRSD----- 279
QY 291 LKHSNLSTEOILDILSLFAGHETSSVAIALAIFLQACPAVEELREHLEIARAKE 350
DB 280 ----PESLOAMKEATELLEFGHETASTATSLVMFLGLNTEYVQKVR--VQEKVE 331
QY 351 LG-----ESELNMDYKMDTQCVINETLRIGNVFELHRAKLDVRYKGYDIPSGMKV 406
DB 332 MCMYTPGKGLSMELLQDLKYTCVIKETLRINPVPVGGFRVALKTFELNGYQIPKGMNYI 391
QY 407 PYSAVHLDNSRYDQPNLFNPMRMOQONGASSSGSSTNGNNMMPGCGPRLCAGSE 466
DB 392 YSICDTHADVAVPKNKEEPQPERFM-----SKIGEDGSR--NYPGGSGRMCVKE 442
QY 467 LAKLEMAVFIHHLVLFKNMELAEDDQ-----PFAFPVDFP 502
DB 443 FAKVLKTLVLELQHCNMLISNGPPTMKTGPIIYVDMLP 483
RESULT 5
C881_MAIZE
ID C881_MAIZE STANDARD: PRT: 519 AA.
AC 043246;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE CYTOCHROME P450 88A1 (EC 1.14.--) (DMARF3 PROTEIN).
GN CYP88A1 OR D3.

```

RESULT 6
CP26_MOUSE STANDARD: PRT: 497 AA.
ID CP26_MOUSE 055127:
AC 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE CYTOCHROME P450 26 (EC 1.14.-.-) (RETINOIC ACID-METABOLIZING
DE CYTOCHROME) (P450RA1) (RETINOIC ACID 4-HYDROXYLASE).
GN CYP26A1 OR CYP26 OR P450RA.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxId=10090;
RN (1)
RP SEQUENCE FROM N.A.
RC STRAIN=C3H.
RX MEDLINE=97392446; PubMed=9250660;
RA Fujii H., Saito T., Kaneko S., Gotcho O., Fujii-Kuriyama Y., Osawa K.,
RA Kato S., Hamada H.;
RT "Metabolic inactivation of retinoic acid by a novel P450
RT differentially expressed in developing mouse embryos.";
RL EMBL J. 16:4163-4173(1997).
RN (2)
RP SEQUENCE FROM N.A.
RX MEDLINE=98113212; PubMed=9442090;
RA Abu-Abied S.S., Beckelt B.R., Chiba H., Chithalen J.V., Jones G.,
RA Metzger D., Champon P., Petkovich M.;
RT "Mouse P450RA1 (CYP26) expression and retinoic acid-inducible retinoic
RT acid metabolism in F9 cells are regulated by retinoic acid receptor
RT gamma and retinoid X receptor alpha.";
RL J. Biol. Chem. 273:2409-2415(1998).
CC -I- FUNCTION: PLAYS A KEY ROLE IN RETINOIC ACID METABOLISM. ACTS ON
CC RETINOIDS, INCLUDING ALL-TRANS-RETINOIC ACID (RA) AND ITS
CC STEREOISOMER 9-CIS-RA. CAPABLE OF BOTH 4-HYDROXYLATION AND 18-
CC HYDROXYLATION. RESPONSIBLE FOR GENERATION OF SEVERAL HYDROXYLATED
CC FORMS OF RA, INCLUDING 4-OH-RA, 4-OXO-RA, AND 18-OH-RA.
CC -I- SUBCELLULAR LOCATION: MEMBRANE-BOUND. ENDOPLASMIC RETICULUM.
CC -I- INDUCTION: BY RETINOIC ACIDS (RA).
CC -I- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: Y12657; CAA73306.1; -.
DR MGD: MGI:1096359; Cyp26.
DR InterPro: IPR001128; -.
DR Pfam: PF00067; P450.1.
DR PRINTS: PR00385; P450.
DR PROSITE: PS00086; CYTOCHROME_P450.1.
KW Oxidoreductase; Monooxygenase; Membrane; Heme; Mitosome;
KW Endoplasmic reticulum.
FT BINDING 442 HEME (POTENTIAL).
FT SEQUENCE 497 AA; 56177 MW; 33B07D7C29134471 CRC64;
Query Match 17.2%; Score 461; DB 1; Length 497;
Best Local Similarity 28.4%; Pred. No. 1.6e-22;
Matches 148; Conservative 78; Mismatches 231; Indels 64; Gaps 14;
OY 10 LPLLPLSL---LSLLFLILK-----RRNRKTRFNLPPKSGMPLGTIGYIK 57
DB 3 LPALLASALCFVPLPLLLFLAALKIMLDYCVSSRRDSCALPLPCTGMFPFEGETL----- 58
OY 58 PYTATTLGDPMQOHNSKRGKIRYSLNLFEEPTIVSADALNPIFLIONEGRLFECSPTSIG 117
DB 59 -QMVLQRRKFFQMKRRKRGKGFITKFLFRPIRVYVGANVRRIILGHRILYSVHNPASVR 117

```

```

QY 118 GILCKMSLVGVGMHDMRSISINFLSHARLRTLLKDVNRHTLVLDMSQO--NSIFS 175
DB 118 TILGAGCLSNLHDSHKKQKVIQAFSREALQCVLV-IAEVSQCLEOMLSCGERGL 176
QY 176 AODAKKTFPLMAKHIMSMPGE-----EETEOLKEEYVTWKGVASPLNPGTAHYKA 231
DB 177 VYPEKRLMFRIARILLKILGCEPAGGEDBOQLEAFEEEMTRNLFSLPIDVPFSGLYRG 236
QY 232 LOSRATILKFERKMEE--RKLDIKEED--OEEDVKTEDBAEMSKSDHVRKOTDDL 286
DB 237 VKARLILHARIEENIRAKIRLQATERPGCKDALQILIEHSWE-----RGERLDQO- 288
QY 287 LGWVLKHSNLSTEOILDLILSLFAGHETSSVAIALAIFLOACPKAVEELREHLEIAR 346
DB 289 ---ALKOS--STE-----LFGGHETSAATSLTLYGLVPHVLQKVRREIKSKGL 335
QY 347 AKKEGESELMWDYKKMDFTQCVINETLRIGNVRFILHRKALKDVRKQIDIPSGMYVL 406
DB 336 LCKSNODKLMETLEOLKTYIGCVIKETRLNPPVPGGFVALKTFELNGYOIPKGMVVI 395
QY 407 FVISAHLNDSRYDQPNLFNPMWMOQNNGASSSGSFSFWNGNMYMPFGGPRICAGSE 466
DB 396 YSICDTHVDALFTNKEEFNDRFIVPHEDASRPS-----FIFPGGLRSCVGEKE 446
QY 467 LAKLEMAVFIHHLVLFKFWMEIAE-----DDQFAFPFVDFP 502
DB 447 FAKILKIFTEVLARHCDWOLNCPPTKTSPTYVPVNDLP 487

RESULT 7
CP26_HUMAN STANDARD; PRT: 497 AA.
AC 043174;
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE CYCLOCHROME P450 26 (EC 1.14.--) (RETINOIC ACID-METABOLIZING
DE CYCLOCHROME) (P450RA1) (RETINOIC ACID 4-HYDROXYLASE).
GN CYP26A1 OR CYP26.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_Taxid=9606;
RN 11
RP SEQUENCE FROM N.A.
RX MEDLINE=97373542; PubMed=9228017;
RA White J.A., Beckett-Jones B., Guo Y.-D., Dilworth F.J., Bonasoro J.,
RA Jones G., Petkovich M.;
RT "cDNA cloning of human retinoic acid-metabolizing enzyme (hP450RA1)
RT identifies a novel family of cytochromes P450.";
RL J. Biol. Chem. 272:18538-18541(1997).
RN 12
RP SEQUENCE FROM N.A.
RX MEDLINE=98380037; PubMed=9716180;
RA Sonneveld E., van den Brink C.E., van der Leede B.M., Schultes R.K.,
RA Pekovic M., van der Burg B., van der Saag P.T.;
RT "Human retinoic acid (RA) 4-hydroxylase (CYP26) is highly specific for
RT all-trans-RA and can be induced through RA receptors in human breast
RT and colon carcinoma cells.";
RL Cell Growth Differ. 9:629-637(1998).
RN 13
RP TISSUE SPECIFICITY.
RX MEDLINE=99045433; PubMed=9826557;
RA Trofimova-Griffin M.E., Juchan M.R.;
RT "Expression of cytochrome P450RA1 (CYP26) in human fetal hepatic and
RT cephalic tissues.";
RL Biochem. Biophys. Res. Commun. 252:487-491(1998).
CC -1- FUNCTION: PLAYS A KEY ROLE IN RETINOIC ACID METABOLISM. ACTS ON
CC RETINOIDS, INCLUDING ALL-TRANS-RETINOIC ACID (RA) AND ITS
CC STEROISOMER 9-CIS-RA. CAPABLE OF BOTH 4-HYDROXYLATION AND 18-
CC HYDROXYLATION. RESPONSIBLE FOR GENERATION OF SEVERAL HYDROXYLATED
CC FORMS OF RA, INCLUDING 4-OH-RA, 4-OXO-RA, AND 18-OH-RA.
CC -1- SUBCELLULAR LOCATION: MEMBRANE-BOUND. ENDOPLASMIC RETICULUM.

```

```

CC -1- TISSUE SPECIFICITY: HIGHEST LEVELS IN ADULT LIVER, HEART,
CC PITUITARY GLAND, ADRENAL GLAND, PLACENTA AND REGIONS OF THE BRAIN.
CC -1- INDUCTION: BY RETINOIC ACIDS (RA).
CC -1- SIMILARITY: BELONGS TO THE CYCLOCHROME P450 FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb.ch/announce/
CC or send an email to license@isb.ch).
CC -----
DR EMBL: AF005418; AAB88881.1; -
DR MIM: 602239; -
DR InterPro: IPR001128; -
DR Pfam: PF00067; P450; 1.
DR PRINTS: PR00385; P450.
DR PROSITE: PS00086; CYCLOCHROME_P450; 1.
KW Oxidoreductase; Monooxygenase; Membrane; Heme; Microsome;
KW Endoplasmic reticulum.
FT BINDING 442 442 HEME (POTENTIAL).
SO SEQUENCE 497 AA; 56162 MW; EAB6B84B24B2EAB3 CRC64;

Query Match 16.3%; Score 438; DB 1; Length 497;
Best Local Similarity 27.0%; Pred. No. 4.9e-21;
Matches 142; Conservative 85; Mismatches 226; Indels 72; Gaps 15;

QY 10 LPALLPSL-----LSLFLILK-----RRNRKTRFNLPKSGMFPGLTGYLK 57
DB 3 LPALLASALCTVPLPLFLFLAIKLMDLYCVSGRDSCLPLPGMGPFFGGLT----- 58
QY 58 PYTATTLGDMOAHVSKGKIRSNLFGPEPTIVSADAGLNRIILONEGLFECSPRSTG 117
DB 59 -QMVLRKRFLLQMKRRKGFYKTHLFGPRVVGADVNRILLGDRIVSVHMPASVR 117
QY 118 GILCKMSLVGVGMHDMRSISINFLSHARL---TILKDVNRHTLVLDMSQO--NS 172
DB 118 TILGAGCLSNLHDSHKKQKVIQAFSREALQCVLV-----LEOMLSCGER 173
QY 173 IFSADDEAKKFTFNLMAKHIMSMP-----GEETEOLKEEYVTWKGVASPLNPGTA 227
DB 174 GLVVPYVKRLMFRIARILLKILGCEPQIAGDQSE-QQLEAVEEMTRNLFSLPIDVPFSG 232
QY 228 YKALQSRATILKFERKMEERKLDIKEED-----OEEDVKTEDBAEMSKSDHVRKOT 282
DB 233 IYRGMKARLILHARIEONIRAKICGLRASEAGQCKDALQILIEHSWE-----RGERL 285
QY 283 DDDDLGWVLKHSNLSTEOILDLILSLFAGHETSSVAIALAIFLOACPKAVEELREHL 342
DB 286 DMQ-----ALKOS--STE-----LFGGHETSAATSLTLYGLVPHVLQKVRREIK 331
QY 343 ETARAKKEGESELMWDYKKMDFTQCVINETLRIGNVRFILHRKALKDVRKQIDIPSG 402
DB 332 SKGLCKSNODKLMETLEOLKTYIGCVIKETRLNPPVPGGFVALKTFELNGYOIPKG 391
QY 403 WKVLPVISAHLNDSRYDQPNLFNPMWMOQNNGASSSGSFSFWNGNMYMPFGGPRIC 462
DB 392 WNVVISICDTHVDALFTNKEEFNDRFIVPHEDASRPS-----FIFPGGLRSC 442
QY 463 AGSELAKLEMAVFIHHLVLFKFWMEIAE-----DDQFAFPFVDFP 502
DB 443 VKKEPAKILKIFTEVLARHCDWOLNCPPTKTSPTYVPVNDLP 487

RESULT 8
YRV2_CABEL STANDARD; PRT: 520 AA.
ID YRV2_CABEL
AC 027514;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)

```

DE PUTATIVE CYTOCHROME P450 CYP13A5 (EC 1.14.-.-).
 GN CYP13A5 OR T10B9.2.
 OS Caenorhabditis elegans.
 CC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
 CC Rhabditidae; Pelodierinae; Caenorhabditis.
 OX NCBI_Taxid=6239;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-BRISTOL N2;
 RA Gardner A.;
 RL Submitted (MAR-1995) to the EMBL/GenBank/DBJ databases.
 CC -1- FUNCTION: CYTOCHROMES P450 ARE A GROUP OF HEME-THIOLATE
 CC MONOOXYGENASES. THEY OXIDIZE A VARIETY OF STRUCTURALLY UNRELATED
 CC COMPOUNDS, INCLUDING STEROIDS, FATTY ACIDS, AND XENOBIOTICS.
 CC -1- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL; Z48717; CAAB8604.1; -
 DR Wormpep; T10B9.2; CE01655.
 DR Interpro; IPR001128; -
 DR Interpro; IPR002402; -
 DR Pfam; PF00067; P450.1.
 DR PRINTS; PR00385; P450.
 DR PRINTS; PR00464; EP45011.
 DR PROSITE; PS00086; CYTOCHROME_P450.1.
 KW Hypothetical protein; Oxidoreductase; Monooxygenase; Heme.
 FT BINDING 464 464 HEME (BY SIMILARITY).
 SO SEQUENCE 520 AA; 59524 MW; 0B7B19E25B7ADFB3 CRC64;

Query Match 12.1%; Score 324.5; DB 1; Length 520;
 Best Local Similarity 23.6%; Pred. No. 9,2e-14;
 Matches 120; Conservative 94; Mismatches 178; Indels 117; Gaps 20;

42 GKSGMPLG-----ETIGLKRYATTTGDEMOQVHVKYKTKYNSLGEPTIYAD 93
 Db 33 GPRGPFEGVTHFQDYNPLK-----LGEWKE-----YGRITGTEGEKTLIYSN 82
 Oy 94 AGLNFIIONGRLFECSYPSIGIGL-----KWSMLVLV-GDMHRMRSISLNFSLHA 147
 Db 83 P---EFVHEVVKQFDNRYGKTNPIQGDPPKKNRAHLVSAQGHKWKRLTLSSFTFSNK 139
 Oy 148 RLRTIL-----LKDVERHTLFLVDSMOONSIFSADKAKFTFNLMKAKHIMSDPGEE 201
 Db 140 NLKRIIMSTVEEIVVELMH---LDDASAKGKAVDLDYQGFETLIDIRIM---GGTE 192
 Oy 202 TEOLKKEVYTFMKV-----VSAPLNLPGTAIYK-----ALQSRATILKF 241
 Db 193 SLMRNPLMPYKGFIFKDRKLPFLVSGIFPIAGTMRFEFMRPPSIOPADIMSTVEKA 252
 Oy 242 IERKMEERKLKE-----EDQ-----EEEVKTEDEAKMSKSDHNRK 279
 Db 253 LNKRIEADAEKAGIEPSGEPQDFIDFLDARANVDFEEESALGAKETIAKVD--K 309
 Oy 280 QRTDDDLGNVLKHSNLTSTEOILDLISLTFAGHETSVAALAIFFLOAPKAVEELRE 339
 Db 310 QLTDEILG-----QLFVFLLAGYDTTALSLSYSLALRRPELOKIOE 354
 Oy 340 EHLIARAKKLGSESLNMDYKKMDFTQVINEFLRLGNVRLH-RKALKDYRYKGYD 398
 Db 355 E-----VDRCPNPEVYTFDOISKLYMECVVKEALRMVPLASIVHNKCKMETNVLGVQ 408
 Oy 399 IPSGMKVLPRVSAVHLDSRY-DQPNLFNPKRMQOONNGASSSGSGSTNGNNTMPGG 457
 Db 409 IEKGTNVQVDTWTTLHYDPKVKVGEADANFRPRWE-----SCDELFLYAKG-GYLPGM 459

Oy 458 GPRCAGSELAKLEMAVFHHLVLEKFNME 486
 Db 460 GPRICIGMRILAMKEMKMLTLIKKTYFE 488

RESULT 9
 YRVA-CAEEL STANDARD; PRT: 518 AA.
 ID YRVA-CAEEL
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 30-MAY-2000 (Rel. 39, Last annotation update)
 DE PUTATIVE CYTOCHROME P450 CYP13A7 (EC 1.14.-.-).
 GN CYP13A7 OR T10B9.10.
 OS Caenorhabditis elegans.
 CC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
 CC Rhabditidae; Pelodierinae; Caenorhabditis.
 OX NCBI_Taxid=6239;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-BRISTOL N2;
 RA Gardner A.;
 RL Submitted (MAR-1995) to the EMBL/GenBank/DBJ databases.
 CC -1- FUNCTION: CYTOCHROMES P450 ARE A GROUP OF HEME-THIOLATE
 CC MONOOXYGENASES. THEY OXIDIZE A VARIETY OF STRUCTURALLY UNRELATED
 CC COMPOUNDS, INCLUDING STEROIDS, FATTY ACIDS, AND XENOBIOTICS.
 CC -1- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL; Z48717; CAAB8609.1; -
 DR Wormpep; T10B9.10; CE01655.
 DR Interpro; IPR001128; -
 DR Interpro; IPR002402; -
 DR Pfam; PF00067; P450.1.
 DR PRINTS; PR00385; P450.
 DR PRINTS; PR00464; EP45011.
 DR PROSITE; PS00086; CYTOCHROME_P450.1.
 KW Hypothetical protein; Oxidoreductase; Monooxygenase; Heme.
 FT BINDING 464 464 HEME (BY SIMILARITY).
 SO SEQUENCE 518 AA; 58999 MW; CC04283EF87B9EA7 CRC64;

Query Match 12.0%; Score 321.5; DB 1; Length 518;
 Best Local Similarity 24.0%; Pred. No. 1.4e-13;
 Matches 120; Conservative 96; Mismatches 184; Indels 101; Gaps 20;

42 GKSGMPLG-----ETIGLKRYATTTGDEMOQVHVKYKTKYNSLGEPTIYAD 93
 Db 33 GPRGPFEGVTHKFTNYENPGLK-----FSEWTKKGYGVYTGTEGEKTLIYSD 82
 Oy 94 AGLNFIIONGRLFECSYPSIGIGL-----KWSMLVLV-GDMHRMRSISLNFSLHA 147
 Db 83 P---EFVHEVVKQFDNRYGKRLTAIQGDPPKKNRVPLVAAQGHKWKRLTLASFTFSNK 139
 Oy 148 RLRTI-----LKDVER-----HTLFLVDSMOONSIFSADKAKFTFNLMKAKH 192
 Db 140 SLKRIIMSTVEEIVVELVSLKASAEKTLMLLEYD-----FTMDIIGMA 187
 Oy 193 MSMDPGEEET-----EOLKKEVYTFMKG-----VSAPLNLPGTAIYK-----ALQSR 236
 Db 188 M-----GQKSLMRNPLMDYKTIFFKGRNNVFMISGIFPVGLALRIAFKPSLDQVAT 243
 Oy 237 TILKFIKMEERKLDLKEEOE---EEVYKTEDEAKM-----SKSDHVRKQRTDDDLG 288
 Db 244 DIQSLERALKMR-LQEDAEKAGIEPSGEPQDFIDFLDARSTVDFEEDAEADQFAKS 302

```

OY 289 WLK-HSMLSTFQIIDLILSLFAGHETSSVAIALAIFLOACPRAVELEHLEIARA 347
D 303 EVLVKDHILTFEIIIGOLFPELLAGYDTALTSSSYLLATHEPICKLOE-----V 356
OY 348 KKEIGESLMDDDKKMFOCVINETLRLGNVARELH-RAKAKVRKRGIDIPSGMKVL 406
D 357 DRECPDPVPTDQSLKLYLECVKKEALRLPLASLVHNRCLKTNTVNLGIEIGTIN 416
OY 407 PVISAVHLDNRY-DQPLFNPWRMOQONNGASSGSGSFSTWGNMPEGGPRCAGS 465
D 417 VDTMSLHHDPRWGDVDVAFEPKRE-----SGDELFFRAGK-GYLPFGMPRICGM 467
OY 466 ELARLEMAVFIHHLVLEKNWE 486
D 468 RLAMMEMKMLTNILKNYTFE 488

RESULT 10
C912-ARATH STANDARD; PRT: 500 AA.
AC 065790: 09S203:
DT 01-OCT-2000 (Rel. 40, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last sequence update)
DE CYTOCHROME P450 91A2 (EC 1.14.-.-).
GN CYP91A2 OR AT4G37430 OR PG17.80.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
OC Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;
OC Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
ON [1]
RP SEQUENCE FROM N. A.
RC STRAIN=CV. COLUMBIA. TISSUE=Seedling;
RX MEDLINE=98281573; PubMed=9620263;
RA Mizutani M., Ward E., Ohta D.;
RT "Cytochrome P450 superfamily in Arabidopsis thaliana: isolation of
RT cDNAs, differential expression, and RFLP mapping of multiple
RT cytochromes P450."
RL Plant Mol. Biol. 37:39-52(1998).
RN [2]
RP SEQUENCE FROM N. A.
RC STRAIN=CV. COLUMBIA;
RX MEDLINE=20083488; PubMed=10617198;
RA Mayer K.F.X., Schueller C., Mambutt R., Murphy G., Volckaert G.,
RA Pohl T., Duesterhoeft A., Stiekema W., Entian K.-D., Terryn N.,
RA Harris B., Ansoorge W., Brandt P., Grivell L., Rieger M.,
RA Weichselgartner M., de Simone V., Obermaier B., Mache R., Mueller M.,
RA Kreis M., Delseny M., Puigdomenech P., Watson M., Schmidheini T.,
RA Reichert B., Portellelle D., Perez-Alonso M., Boutry M., Bancroft I.,
RA Vos P., Hohenseil J., Zimmermann W., Medler H., Ridley P.,
RA Langham S.-A., McCullagh B., Bilham L., Robben J.,
RA Van der Schueren J., Grymonprez B., Chung Y.-J., Vandenbussche F.,
RA Braeken M., Welfjens I., Voet M., Bastlaens I., Aert R., Defoor E.,
RA Holtzenger T., Bothe G., Ramsperger U., Hilbert H., Braun M.,
RA Holzer E., Brandt A., Peters S., van Staveren M., Dikse W.,
RA Moolman P., Klein lankhorst R., Rose M., Hauf J., Koeter P.,
RA Benelaiser S., Hempel S., Feldpausch M., Lambrecht S., Van den Daele H.,
RA De Keyser A., Blyssnaert C., Gielen J., Villarroel R., De Clercq R.,
RA Van Montagu M., Rogers J., Cronin A., Quail M., Bray-Allen S.,
RA Clark L., Doggett J., Hall S., Kay M., Leonard N., Molay K., Mayes R.,
RA Petiet A., Rajandream M.-A., Lyne M., Benes V., Reichmann S.,
RA Borkova D., Blocker H., Scharfe M., Grimm M., Loehner T.-H.,
RA Dose S., de Haan M., Maarse A., Schaefer M., Mueller-Auer S.,
RA Gabel C., Fuchs M., Fatmann B., Granderath K., Dauner D., Herzi A.,
RA Neumann S., Argitlou A., Vitale D., Liguori R., Piravandi E.,
RA Massenot O., Quigley F., Glabaud G., Muendlein A., Felber R.,
RA Schnabl S., Hiller R., Schmidt W., Lecharny A., Aubourg S.,
RA Cheifod F., Cooke R., Berger C., Monfort A., Casacuberta E.,
RA Gibbons T., Weber N., Vandenbol M., Bauges M., Terol J., Torres A.,
RA Perez-Perez A., Purnelle B., Bent E., Johnson S., Tacon D., Jesse T.,
RA Heijnen L., Schwarz S., Scholler P., Heber S., Francis P., Bielke C.,
RA Frishman D., Haase D., Lemcke K., Mewes H.-W., Stocker S.,

```

```

RA Zaccaria P., Beyan M., Wilson R.R., de la Bastide M., Habermann K.,
RA Pannell L., Dedha N., Gnoj L., Schütz K., Huang E., Spiegel L.,
RA Sakon M., Murray J., Sheet P., Cordes M., Abu-Threideh J.,
RA Stoneking T., Kalicki J., Graves T., Harmon G., Edwards J.,
RA Latreille P., Courtney L., Cloud J., Abbott A., Scott K., Johnson D.,
RA Mix P., Bentley D., Fulton B., Miller N., Greco T., Kemp K.,
RA Krimer J., Fulton L., Mardis E., Dante N., Pepin K., Hillier L.,
RA Nelson J., Spith J., Ryan E., Andrews S., Geisel C., Layman D.,
RA Du H., Ali J., Berghoff A., Jones K., Drone K., Cotton M., Joshi C.,
RA Antonou B., Zidanic M., Strong C., Sun H., Lamar B., Jordan C.,
RA Ma P., Zhong J., Preston R., Vil D., Shekher M., Matero A., Shah R.,
RA Swaby I.K., O'Shaughnessy A., Rodriguez M., Hoffman J., Tili S.,
RA Granat S., Shohdy N., Hasegawa A., Hamed A., Lodhi M., Johnson A.,
RA Chen E., Marra M., Martienssen R., McCombie W.R.;
RT "Sequence and analysis of chromosome 4 of the plant Arabidopsis
RT thaliana."
RL Nature 402:769-777(1999).
CC -i- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb.ch/announce/
CC or send an email to license@isb.ch).
CC
DR EMBL: D78607; BAA28539.1;
DR EMBL: AL035601; CAB38210.1;
DR EMBL: AL061591; CAB80408.1;
DR Mendel: 29894; Arath:1113.29894.
DR InterPro: IPR001128;
DR Pfam: PF00067; P450.1.
DR PROSITE: PS00086; CYTOCHROME_P450.1.
KW Oxidoreductase; Monooxygenase; Membrane; Heme.
FT BINDING 438 438 HEME (BY SIMILARITY).
FT CONFLICT 106 106 A -> V (IN REF. 1).
FT CONFLICT 127 127 I -> M (IN REF. 1).
FT CONFLICT 140 140 N -> I (IN REF. 1).
FT CONFLICT 454 454 S -> T (IN REF. 1).
SQ SEQUENCE 500 AA; 57555 MW; 0FB453D2070EA2EA CRC64;

Query Match 11.8%; Score 316; DB 1; Length 500;
Best Local Similarity 25.2%; Pred. No. 3.1e-13;
Matches 135; Conservative 83; Mismatches 199; Indels 118; Gaps 22;

OY 9 LPLLLPSLLSLFLILKLR-RNRKTRFNLPPGKSGMPLGFTIGYKPYATTLGD 66
D 1 MLYFLLP-----LLFVYSYFLYLSKQRFNLPFGPPSPRPVGH-LHLMKP-----PIHR 50
OY 67 FMOQHVSKYKTKYSNLSFGDEPTV-----SADAGINRFILONEGRLEFCSYPRSTG 117
D 51 LLQRRSYNDYGPFL-SLRGSRVRVITSPSLAOESFTGNDIVLSRPLQLTAKVAVVNH 109
OY 118 GLTKWMSLVLVGDHHRMRST-SLNPISHARLRTI-LTKOVENHTLEVDLSMOONSIF 174
D 110 TTVG-----TAPYGDWRNRLRICSOEILSHRLINFOHNRKDELLRMLRLSRVYOTS- 163
OY 175 SAODEAKKRTF-----NLMAKHMSMDPC-----EEETDLKKKEYVFMKGV 216
D 164 ---NSNDFTHELRLPLSDLFNNIVRVYKRYGDDVNNKEADLPKK----- 211
OY 217 VSAPLNLTGTAVKALQSRAT-----LKTETKMEERKLDIEEDOEVEVYTEDEA 269
D 212 -----LVYDIDAMYSGANHSADYLPILKLGKPKFEVYKAI-----KSMDDI 253
OY 270 ESKSDHVRKORTDDLDLGWVLKHSNLSTFOLD-----LILSLFAGHETSSVAIALAI 324
D 254 LQRLIDECRRDKRGSTVNNHLSLSDQOPEYTTVYIINGLMSMMLAGTETSAYVLEMAN 313
OY 325 FFLQACPRAVELEHLEIARAKKEIGESSEL-NMDDYKKMDFQCVINETLRLGNVARE 383

```

Db 314 ANLRNPEVLEKARSE-----IDKIGKRLIDESDIAVLPLQNVSETFLRPVAF 367
 Oy 384 L-HRRALKDVYKGYDIPSGMKVLPVISAHLNDSRYDOPNLFNPMRQOONNGASSSSG 442
 Db 368 LIPSPIDMDMIGGVDYRDTIVWAMAIHRDPEIMEPEKFPNDRK-----NDGC----- 419
 Oy 443 GSFSTGWNVY-----MFGGGRPLCAGSELAKLEMAVFIHHLVLKFMWELAEDDO 492
 Db 420 -----GSDYVYVKMPFGNGRRTCPGALGORIVTALGSLIOCFEWEVNGKEE 468

RESULT 11
 CP32_RAT STANDARD: PRT: 504 AA.
 ID CP32_RAT 064672: 064629;
 AC P05183; 064672: 064629;
 DT 13-AUG-1987 (Rel. 05, Created)
 DT 15-JUL-1999 (Rel. 38, Last sequence update)
 DT 15-JUL-1999 (Rel. 38, Last annotation update)
 DE CYTOCHROME P450 3A2 (EC 1.14.14.1) (CYP1A2) (P450-PCN2) (P450/6-BETA-7) (TESTOSTERONE 6-BETA-HYDROXYLASE).
 GN CYP3A2 OR CYP3A-2.
 OS Rattus norvegicus (Rat).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus. NCBI_TaxId=10116;
 RX MEDLINE-SPRAGUE-DANLEY: TISSUE=Liver;
 RA Miyata M., Nagata K., Yamazoe Y., Kato R.;
 RL "A gene structure of testosterone 6 beta-hydroxylase (P45011IA)."; Biochem. Biophys. Res. Commun. 177:68-73(1991).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN=SPRAGUE-DANLEY; TISSUE=Liver;
 RX MEDLINE-91254339; PubMed=2043144;
 RA Miyata M., Nagata K., Yamazoe Y., Kato R.;
 RL "Structure of a gene and cDNA of a major constitutive form of testosterone 6 beta-hydroxylase (P450/6 beta A) encoding CYP3A2: comparison of the cDNA with P450PCN2." Arch. Biochem. Biophys. 314:351-359(1994).
 RL Arch. Biochem. Biophys. 314:351-359(1994).
 CC -1- FUNCTION: CYTOCHROMES P450 ARE A GROUP OF HEME-THIOLATE MONOOXYGENASES. IN LIVER MICROSOMES, THIS ENZYME IS INVOLVED IN AN NADPH-DEPENDENT ELECTRON TRANSPORT PATHWAY. IT OXIDIZES A VARIETY OF STRUCTURALLY UNRELATED COMPOUNDS, INCLUDING STEROIDS, FATTY ACIDS, AND XENOBIOTICS.
 CC -1- CATALYTIC ACTIVITY: RH + REDUCED FLAVOPROTEIN + O(2) -> ROH + OXIDIZED FLAVOPROTEIN + H(2)O.
 CC -1- SUBCELLULAR LOCATION: MEMBRANE-BOUND. ENDOPLASMIC RETICULUM.
 CC -1- INDUCTION: BY PREGNENOLONE 16-ALPHA-CARBOXYTRILE (PCN2).
 CC -1- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
 CC
 DR EMBL: M13646; AAA41051.1; -
 DR EMBL: X79319; CA55887.1; -
 DR EMBL: X79320; CA55888.1; -
 DR EMBL: U09742; AA82168.1; -
 DR EMBL: U09734; AAB60492.1; -
 DR EMBL: U09735; AAB60492.1; JOINED.

DR EMBL: U09726; AAB60492.1; JOINED.
 DR EMBL: U09727; AAB60492.1; JOINED.
 DR EMBL: U09728; AAB60492.1; JOINED.
 DR EMBL: U09729; AAB60492.1; JOINED.
 DR EMBL: U09730; AAB60492.1; JOINED.
 DR EMBL: U09731; AAB60492.1; JOINED.
 DR EMBL: U09732; AAB60492.1; JOINED.
 DR EMBL: U09733; AAB60492.1; JOINED.
 DR PIR: A25222; A25222.
 DR HSSP: P14779; IFAG.
 DR InterPro: IPR001128; -
 DR InterPro: IPR002397; -
 DR InterPro: IPR002401; -
 DR InterPro: IPR002402; -
 DR InterPro: IPR002403; -
 DR Pfam: PF00067; P450.1.
 DR PRINTS: PR00359; BP450.
 DR PRINTS: PR00385; P450.
 DR PRINTS: PR00463; EP450I.
 DR PRINTS: PR00464; EP450II.
 DR PRINTS: PR00465; EP450IV.
 DR PROSITE: PS00086; CYTOCHROME P450: 1.
 KW Oxidoreductase; Monooxygenase; Electron transport; Membrane; Heme;
 KW Microsome; Endoplasmic reticulum.
 FT BINDING 443 443 HEME.
 FT CONFLICT 56 57 GL -> AV (IN CA55887).
 FT CONFLICT 429 429 D -> H (IN REF. 1).
 FT CONFLICT 445 445 G -> D (IN REF. 1).
 SQ SEQUENCE 504 AA; 57731 MW; C2C84AF736035AB2 CRC64;

Query Match 11.7%; Score 312.5; DB 1; Length 504;
 Best Local Similarity 25.5%; Pred. No. 5.2e-13;
 Matches 137; Conservative 96; Mismatches 211; Indels 93; Gaps 24;

Oy 9 LPLLLPSLLSLFLILKR--RNKTRNLP--PEKSGPPLGELIGLKYTTATL 64
 Db 3 LLSALTLETWLVLLVILVLRGLGTHRGIGKKGIDGPKPLPLGTVLNYKK-----GL 57
 Oy 65 GDFMQHVKYKGIYRSNLFGPEPTVSA--DAGLNRFLONEGRLFEFC---SYPRSIG- 117
 Db 58 GRFMECKYKGIKTI--GLFDGQTPVPAIMDEMIKNVL-----VKECFSVFTNRDEGP 110
 Oy 118 -GILKSMVLVGDHMDMSISLNFLSHARLRTILKDYERHTLFYLSWQONS---- 172
 Db 111 VGINGKAVSVAKDEMKRYRALLSPFTFS-GRLE-MPEIIQGDILVKYKQAEATGK 168
 Oy 173 -----IFSA--QDEAKFTFNMAKHIMS-MDPGEETEDLKR--EYVTFMKGVASAPL 221
 Db 169 PVTMKVFGAYSMDVITSTSGVAVDSLNPKDPFEVETKTLRFDFEDPLFLSVLPFP 228
 Oy 222 NLPGTAVKHAL-----QSRATILKFERKMEERKLDKEEDDEEVEKTEDEAKMSK 274
 Db 229 LTP---IYEMLNTMFPKDSIAFQKFEVHR-IKETRLDSKKHR-----VDPLQLMLN 277
 Oy 275 DHRVKORTDDLLGVMLKHSNLTQIIDLILSLFAGHETSVAIALAIFLOACPKAV 334
 Db 278 AH---NNSKDE-----VSHKRLSDVEIIAOSVIFIFAGYETSTSLSVLFLATHPDIO 329
 Oy 335 ELREHEHLEIARAKKEIGESELNMDYKKMDQTCVNETRLGAVVFLHRAKLQVRY 394
 Db 330 KRIQEE-IDGALPSK-----APPTVDIWMEMYLDVNETRLXPLDIGNRLERVCCKDIEL 384
 Oy 395 KGVIIPSGMKVLPVISAHLNDSRYDOPNLFNPMRQOONNGASSSGSSTGNNYMP 454
 Db 385 DGLIPKGSVVTIPTVLAHNDPQHPMPKEEFHPRFSEKNGKSIDP-----YYVLP 435
 Oy 455 FGCGPRLCAGSELAKLEMAVFIHHLVLKFMWELAEDDOPFAFPVDPNGLPVRSR 511
 Db 436 FGNGPRNCIGMRFAIMNKKALITVLQNFSPQCKETQ-----IPLKLSR 480

RESULT 12


```

Db 140 SLRKIMSTVEETVEELMHLDDEASAKGAVLDLDYQE-----FTLDIGRIA 187
OY 193 MSMDEGEET-----EOLKKEVTFPMKG-----VVSAPLNPGT-----AYHKAL 232
Db 188 M-----GQIESLMEFRNPMPLPKYKEIFKKGKMPFLIAGVFPILGILMRLMKFPKFSFAF 243
OY 233 OSRATILAFIERKMEERKLDIRE-----EDQEEVEKTEDEAEMSKSDHYRKORTDDL 286
Db 244 GIMNTMEKALNRLEQRAADKKAGIEPSEPODFIDFLDARANY--DFIEEST----- 296
OY 287 LGM-----VLK-HSNLSTQIIDLILSLFAGHETSVAIALAIFLQCPRAVEELREEH 341
Db 297 LGFAKSEVLAKVDKHLTFPEIIGQLFVFLIAGVDTALSLSSYLLATHPEIOKKLOEE- 355
OY 342 LEIARAKKELGSELMNDYKMKDFTOCVINEITRLGNVRFILH-RKALKDVRKGYDIP 400
Db 356 -----VDECEPDPEYTFQIOLSKLMECVKALMRYPLASLVHNRKCKMTNVLGVEID 410
OY 401 SGWKVLPVISAVALDINSRY-DQPLNFNPMRQOONNGASSSGSFSFGWNNYMPFGGCP 459
Db 411 EGTNVQVDTWTLHDPKQWGDASEFKPERME-----TGDELFFYAKG-GYLPGMGMP 461
OY 460 RLCAGSELAKLEMAVFIHHLVLEKEMLEAEDQ 492
Db 462 RICIGMRLAMMEKILLTHILKLYFTSTETE 494

```

RESULT 14

CP36_MOUSE STANDARD: PRT: 501 AA.

AC 054750;

DT 15-DEC-1998 (Rel. 37, Created)

DT 15-DEC-1998 (Rel. 37, Last sequence update)

DT 01-OCT-2000 (Rel. 40, Last annotation update)

DE CYTOCHROME P450 2J6 (EC 1.14.14.1) (CYPIIb) (ARACHIDONIC ACID

DEPOXYGENASE).

GN CYP2J6.

OS Mus musculus (Mouse).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

OX NCBI_TaxID=10090;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=C57BL/6 X CBA; TISSUE=Liver;

RX MEDLINE=98234557; PubMed=9570962;

RA Ma J., Ramachandran S., Fiedorek F.T. Jr., Zeldin D.C.;

"Mapping of the CYP2J cytochrome P450 genes to human chromosome 1 and

mouse chromosome 4.";

RT Genomics 49:152-155(1998).

RL -1- CATALYTIC ACTIVITY: RH + REDUCED FLAVOPROTEIN + O(2) = ROH +

OXIDIZED FLAVOPROTEIN + H(2)O.

CC -1- SUBCELLULAR LOCATION: MEMBRANE-BOUND. ENDOPLASMIC RETICULUM.

CC -1- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

between the Swiss Institute of Bioinformatics and the EMBL outstation -

the European Bioinformatics Institute. There are no restrictions on its

use by non-profit institutions as long as its content is in no way

modified and this statement is not removed. Usage by and for commercial

entities requires a license agreement (See <http://www.isb.ch/announce/>

or send an email to license@isb-sib.ch).

CC EMBL: U62295; AAB87636.1; -

DR MGD; MGI:1270148; Cyp2j6.

DR InterPro: IPR001128; -

DR InterPro: IPR002401; -

DR Pfam: PF00067; P450; 1.

DR PRINTS: PR00385; P450.

DR PRINTS: PR00463; EP4501.

DR PROSITE: PS00086; CYTOCHROME_P450; 1.

DR Oxidoreductase; Monooxygenase; Electron transport; Membrane; Heme;

KW Microsome; Endoplasmic reticulum.

FT BINDING 447 447 HEME (BY SIMILARITY).

SQ SEQUENCE 501 AA; 57820 MW; 4E5A26B35D831E37 CRC64;

Query Match 11.3%; Score 303; DB 1; Length 501;
Best Local Similarity 23.2%; Pred. No. 2,1e-12;
Matches 126; Conservative 98; Mismatches 204; Indels 116; Gaps 21;

```

OY 9 LPLLLPSLLSLFLILKRRNKRTPNLPKGSMPPIGFTIGYLKPTATLUGDFM 68
Db 16 LHPRLVLAATVFLILLADYFKNRKPK---NPPGWMGLPFGNIFOL-----DFG 62
OY 69 QQWVS-----KYGRYNSN-----FGEPTVSADAGLREPILONEGRLECSYPR 114
Db 63 QPHLSIOPLVKYGNIFFSLNGDITSVITGILPKLKALTOGEONIMNR-----PL 113
OY 115 SI--GGILGKSMVLVVDMDHMRSLNLSNARL-RTLLKDVENHTLFLVDSMOON 171
Db 114 SVMQERISNKNGLIFSSQIWKQORFALMTLRNGLGKSLSEERMQDEASHLVEAIR-- 171
OY 172 SIFSADAKKFT-----FNLMAKHMSMDGE-----EETOLKKEVTFPMKGCV 217
Db 172 -----EEGKFPNPPFISNNVSNITCSVTGERPDYDSRQEMLRLLDEVYMLLETTMI 226
OY 218 SAPLN-----LPGTAYHRALOSRAITLKFIERKMEERKLDIREDOEEVEKTEDEA 269
Db 227 SOLVNIFFPMIKKIYIPG-SHOKVFRWMEKLFVSCMIDHR-----KQWNPDEPRDFIDA 280
OY 270 ---EKSQSDHVRKORTDDLLGWLKHSNLSSTEQIIDLILSLFAGHETSSVAIALAIF 326
Db 281 FLKEWTK-----YPEKTSFNEENLICSTLDFAGFETSTLLRMALLY 325
OY 327 LQACPAVEELREHLEIARAKKELGSELMWMDYKKKDDFOCVYNEITRLGNVRF-LH 385
Db 326 MALPPEQEKVQAEIDRYIGQR-----AARLADRESMPYTNNAVYHEHRRGNITLPLNP 380
OY 386 RKALKDVRYKGYDIPSGMKVLPVISAVALDINSRYDQPLNFNPMRQOONNGASSSGSFS 445
Db 381 REVAMDTNLNFFHLRPLKGMVLTNLTLALHDPKEMATPVFNPHELE-----NQGF 431
OY 446 STWGNINYPFGGCPRLCAGSELAKLEMAVFIHHLVLEK-FWMELEDDQPAFPVDPN 503
Db 432 KK-RESFLPFSMGKRAIGEOLARSELPIFTSLMKQKTFNPINELSP-----KFRN 484
OY 504 GLPI 507
Db 485 GLTL 488

```

RESULT 15

CPA8_MESAU STANDARD: PRT: 494 AA.

AC P24454;

DT 01-MAR-1992 (Rel. 21, Created)

DT 01-MAR-1992 (Rel. 21, Last sequence update)

DE CYTOCHROME P450 2A8 (EC 1.14.14.1) (CYPIIa8) (P450-MC1) (P450-AFB).

GN CYP2A8.

OS Mesocricetus auratus (Golden hamster).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae;

OC Mesocricetus.

OX NCBI_TaxID=10036;

RN [1]

RP SEQUENCE FROM N.A.

RC TISSUE=Liver;

RX MEDLINE=9112759; PubMed=2275554;

RA Lai T.S., Chiang J.Y.L.;

"Cloning and characterization of two major 3-methylcholanthrene

inducible hamster liver cytochrome P450s.";

RT Arch. Biochem. Biophys. 283:429-439(1990).

RN [2]

RP SEQUENCE FROM N.A.

```

RC TISSUE=Liver;
RX FUKUTERA M., Nagata K., Mizokami K., Yamazoe Y., Takanaka A., Kato R.;
RA "Complete cDNA sequence of a major 3-methylcholanthrene-inducible
RT cytochrome P-450 isozyme (P-450AEB) of Syrian hamsters with high
RL Biochem. Biophys. Res. Commun. 162:265-272(1989).
CC -1- FUNCTION: HIGHLY ACTIVE IN 7-ETHOXYCOUMARIN O-DEETHYLATION, AND
CC BENZHEPTAMINE N-DEETHYLATION; MODERATELY ACTIVE IN TESTOSTERONE
CC 7-ALPHA-HYDROXYLATION, ETHYLORPHINE N-DEETHYLATION, P-
CC NITROANISOLE O-DEETHYLATION, AND ONLY SLIGHTLY ACTIVE IN
CC BENZOPYRENE 3-HYDROXYLATION, 7-ETHOXYRESORUFIN O-DEETHYLATION,
CC TESTOSTERONE 2-ALPHA-HYDROXYLATION AND TESTOSTERONE 17-OXIDATION.
CC COMPETENT IN THE METABOLIC ACTIVATION OF AFLATOXIN B1.
CC -1- CATALYTIC ACTIVITY: RH + REDUCED FLAVOPROTEIN + O(2) = ROH +
CC OXIDIZED FLAVOPROTEIN + H(2)O.
CC -1- SUBCELLULAR LOCATION: MEMBRANE-BOUND. ENDOPLASMIC RETICULUM.
CC -1- TISSUE SPECIFICITY: LIVER.
CC -1- INDUCTION: BY 3-METHYLCHOLANTHRENE (3MC).
CC -1- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M63788; AAA37068.1; -
DR EMBL; M27906; AAA37084.1; -
DR PIR; A33293;
DR PIR; S13884; S13884.
DR InterPro; IPR001128; -
DR InterPro; IPR002401; -
DR Pfam; PF00067; P450.1.
DR PRINTS; PR00385; P450.1.
DR PRINTS; PR00463; EP4501.
DR PROSITE; PS00086; CYTOCHROME_P450.1.
DR Oxidoreductase; Monooxygenase; Electron transport; Membrane; Heme;
KW Microsome; Endoplasmic reticulum.
FT BINDING 439 439 HEME (BY SIMILARITY).
FT CONFLICT 417 417 S -> L (IN REF. 2).
SO SEQUENCE 494 AA; 57387 MW; 16816672AAAB112B CRC64;

Query Match 11.3%; Score 302; DB 1; Length 494;
Best Local Similarity 24.2%; Pred. No. 2,4e-12;
Matches 128; Conservative 85; Mismatches 228; Indels 88; Gaps 20;

QY 12 LLLPSSLILFLILKRRNRKTRFNLPPKSGWPIGFTIGYIKPYTATTLGDFMOOH 71
DB 6 MLVAVVLTCLSMVIMISVYRQRRLKMPGPTPLPFTGNFL---ELDTREYDCLSKM 61
QY 72 VSKYGIYRSMIFGEPTIV--SADAGLNRFILONGRLEFECSSYPSISGILGKMSMLVLY 129
DB 62 RERYGPVTTIHLGPRPAVLMGYDAVKAALIDQAE-ELSDRGEQAFDFWFRKGIVVSS 120
QY 130 GDMHRMRSISLNFSLH-----ARLRTI-----LLKDVRRHTL---FVLDWMOON 171
DB 121 GERAKQLRFSIATLRDGFGRGIEERTIETSFLLQALRDNTGATIDPTFYMSRTVSN 180
QY 172 SIFSA-----QDEAKKTFNLMKAIMSDGEEETQALKREYTFMKGVASAPLNLPG 225
DB 181 VISSIVFENREYDKEF-LSLGIMRSFQFMSTSTQLEFEMFYSVK-----HLPG 232
QY 226 TAYKALQSRATILKFERKME--RKLDIKEDD--EEEEKTFDEAMSKSD-HVRKQ 280
DB 233 -CQHAYKEMOGLEDFLARKVEENORTLDPSNPRDFISFLIRMOEKKNPRTQFHR-- 289
QY 281 RTDDDLGWLKHSNLSLEQILDLILSLFAGHETSVAIALATFFLOACPKAVAEELREE 340
DB 290 -----NLMTLTNLNFFAGTEIVSTTTTRYGFLLLMKYPHIAKMHHE 330

```

```

QY 341 HLEIARAKKEIGES-ELAWDDYKKMDFTCYINETLRIGNVYRF-LHRKALKDVRKGYD 398
DB 331 IDVY-----IGRNQPKYEDHLKMPYEAVIETIQRPVDVVPGLPSTTKDKIFDFL 384
QY 399 IPGKVLPIVISAVALDNSRYDQPNLFNPMRWQDQONNGASSSGSFSFGWGNVMPFGCG 458
DB 385 IPKGTDFVPLSSVLKDKPKFSNPNDFNPQH-----SDDKQGFKK-SNAFMPESVG 435
QY 459 PRLCAGSELAEVAVFTIHLVLFKNWELAE----DDQPAFPVDFP 502
DB 436 KRYCGESLAKMELFIFFTLMQNFCEKSPQAPQDIDVTPOYFSFAALP 484

```

Search completed: October 17, 2001, 22:14:09
Job time: 363 sec


```

Db 121 GKWSMLVGVGMHHRDMRSISLNFLSHARLRTLLKDVBRHTLFVLDMSQNSIFSADDEA 180
QY 181 KKFTFNMAKIMMSMDPGEETEOLKKEYVTFMKGVSAPLNPGTAVHKAQSRATILK 240
Db 181 KKFTFNMAKIMMSMDPGEETEOLKKEYVTFMKGVSAPLNPGTAVHKAQSRATILK 240
QY 241 FIERKMERKLDIKEEDOEVEEVKTEDEAEKSDHVRKQRTDDLLGWLKHSNLSTEQ 300
Db 241 FIERKMERKLDIKEEDOEVEEVKTEDEAEKSDHVRKQRTDDLLGWLKHSNLSTEQ 300
QY 301 ILDLILSLFAGHETSSVAIALAIFLQACPKAVEELREHELEIARAKKEGSELMND 360
Db 301 ILDLILSLFAGHETSSVAIALAIFLQACPKAVEELREHELEIARAKKEGSELMND 360
QY 361 YKKMDFTOCVINETLRIGNVVRFHLRAKAKDVRKGYDIPSGMKVLPVISAHLNDSRYD 420
Db 361 YKKMDFTOCVINETLRIGNVVRFHLRAKAKDVRKGYDIPSGMKVLPVISAHLNDSRYD 420
QY 421 QPNLFNPMRMOQONNGASSSGSGSFTWGNVMPFGGPRLCAGSELAKLEMAVFIHHLV 480
Db 421 QPNLFNPMRMOQONNGASSSGSGSFTWGNVMPFGGPRLCAGSELAKLEMAVFIHHLV 480
QY 481 LKFNWELAEDDQPPAFPFVDFPENGPIRVSRIL 513
Db 481 LKFNWELAEDDQPPAFPFVDFPENGPIRVSRIL 513

RESULT 2
Q9SCQ9 PRELIMINARY; PRT; 513 AA.
AC Q9SCQ9;
DT 01-MAY-2000 (TREMblrel. 13, Created)
DT 01-MAY-2000 (TREMblrel. 13, Last sequence update)
DT 01-MAR-2001 (TREMblrel. 16, Last annotation update)
DE STEROD 22-ALPHA-HYDROXYLASE (DWF4).
GN T345.40.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
OC Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;
OC Brassicaceae; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RA Bioecker H., Mewes H.W., Lemcke K., Mayer K.F.X., Quetier F.,
RA Salanoubat M.;
RL Submitted (NOV-1999) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA EU Arabidopsis sequencing project;
RL Submitted (DEC-1999) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.
CC EMBL: AL132979; CAB62435.1; -.
DR InterPro: IPR001128; -.
DR Pfam: PF00067; P450.1.
DR PRINTS: PR00385; P450.
DR PROSITE: PS00086; CYTOCHROME_P450; UNKNOWN_1.
KM Heme; Monooxygenase; Oxidoreductase.
SQ SEQUENCE 513 AA; 58867 MW; B1639BD9AD5D7C93 CRC64;

Query Match 99.9%; Score 2677; DB 10; Length 513;
Best Local Similarity 99.8%; Pred. No. 1.2e-182;
Matches 512; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 121 GKWSMLVGVGMHHRDMRSISLNFLSHARLRTLLKDVBRHTLFVLDMSQNSIFSADDEA 180
Db 121 GKWSMLVGVGMHHRDMRSISLNFLSHARLRTLLKDVBRHTLFVLDMSQNSIFSADDEA 180
QY 181 KKFTFNMAKIMMSMDPGEETEOLKKEYVTFMKGVSAPLNPGTAVHKAQSRATILK 240
Db 181 KKFTFNMAKIMMSMDPGEETEOLKKEYVTFMKGVSAPLNPGTAVHKAQSRATILK 240
QY 241 FIERKMERKLDIKEEDOEVEEVKTEDEAEKSDHVRKQRTDDLLGWLKHSNLSTEQ 300
Db 241 FIERKMERKLDIKEEDOEVEEVKTEDEAEKSDHVRKQRTDDLLGWLKHSNLSTEQ 300
QY 301 ILDLILSLFAGHETSSVAIALAIFLQACPKAVEELREHELEIARAKKEGSELMND 360
Db 301 ILDLILSLFAGHETSSVAIALAIFLQACPKAVEELREHELEIARAKKEGSELMND 360
QY 361 YKKMDFTOCVINETLRIGNVVRFHLRAKAKDVRKGYDIPSGMKVLPVISAHLNDSRYD 420
Db 361 YKKMDFTOCVINETLRIGNVVRFHLRAKAKDVRKGYDIPSGMKVLPVISAHLNDSRYD 420
QY 421 QPNLFNPMRMOQONNGASSSGSGSFTWGNVMPFGGPRLCAGSELAKLEMAVFIHHLV 480
Db 421 QPNLFNPMRMOQONNGASSSGSGSFTWGNVMPFGGPRLCAGSELAKLEMAVFIHHLV 480
QY 481 LKFNWELAEDDQPPAFPFVDFPENGPIRVSRIL 513
Db 481 LKFNWELAEDDQPPAFPFVDFPENGPIRVSRIL 513

RESULT 3
Q9LKH7 PRELIMINARY; PRT; 474 AA.
AC Q9LKH7;
DT 01-OCT-2000 (TREMblrel. 15, Created)
DT 01-OCT-2000 (TREMblrel. 15, Last sequence update)
DT 01-MAR-2001 (TREMblrel. 16, Last annotation update)
DE CYTOCHROME P450.
GN CIPCYP.
OS Phaseolus aureus (Mung bean) (Vigna radiata).
OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
OC Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids I;
OC Fabales; Fabaceae; Papilionoideae; Vigna.
OX NCBI_TaxID=3916;
RN [1]
RP SEQUENCE FROM N.A.
RA Yang M.T., Chen Y.M.;
RA Yang M.T., Chen Y.M.;
RL Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.
CC -1- CATALYTIC ACTIVITY: RH + REDUCED FLAVOPROTEIN + O(2) -> ROH +
CC OXIDIZED FLAVOPROTEIN + H(2)O (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: MEMBRANE-BOUND. ENDOPLASMIC RETICULUM (BY
CC SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.
CC EMBL: AF279252; AF89209.1; -.
DR InterPro: IPR001128; -.
DR Pfam: PF00067; P450.2.
DR PRINTS: PR00385; P450.
DR PROSITE: PS00086; CYTOCHROME_P450; UNKNOWN_1.
KM Electron transport; Endoplasmic reticulum; Heme; Membrane; Microsome;
KM Monooxygenase; Oxidoreductase.
SQ SEQUENCE 474 AA; 54037 MW; A2542A809C5BACD CRC64;

Query Match 35.5%; Score 951; DB 10; Length 474;
Best Local Similarity 40.1%; Pred. No. 8.2e-60;
Matches 203; Conservative 97; Mismatches 162; Indels 44; Gaps 10;

```

```

QY 1 MEFEHHTLPLLLPSLISLLFLILKRRNRKTRNLPFGKSGWPLFGTTGYLKPRT 60
Db 1 MEFEHHTLPLLLPSLISLLFLILKRRNRKTRNLPFGKSGWPLFGTTGYLKPRT 60
QY 61 ATTIGDMQOVSKYKTKYRSNLFGEPTIYSADAGLNRTIQLNGRGLFECYPRSIGIL 120
Db 61 ATTIGDMQOVSKYKTKYRSNLFGEPTIYSADAGLNRTIQLNGRGLFECYPRSIGIL 120

```

```

QY 12 LLLPSLISLLF-----LILKRRNRKTRNLPFGKSGWPLFGTTGYLKPRTATTLGD 66
Db 1 MVSILPTL-LILFAASAIAIFLHRAFSRRKFRILPPGSGYGLPIGTQLTQLISAVKSSNPP 58

```


DR PROSITE: PS00037; MYB_1; UNKNOWN_1.
 KW Electron transport; Endoplasmic reticulum; Heme; Hypothetical protein;
 KM Membrane; Microsome; Monooxygenase; Oxidoreductase.
 SQ SEQUENCE 524 AA; 59389 MW; 550578908BDDF272 CRC64;

Query Match 32.1%; Score 864; DB 10; Length 524;
 Best Local Similarity 35.4%; Pred. No. 1,5e-53;
 Matches 186; Conservative 96; Mismatches 165; Indels 78; Gaps 10;

QY 24 FLIL-----LKRNRKRTFN-----LPPGSGMFLGFTGYL--- 56
 DB 30 FLVLTAGILRPWILFRILNRKSTKDGDEEDNEKKKGIMGSLGVETLNFILGCG 89
 QY 57 ---KPYATTLGDMQOQVSKYKIRSNLFGPEPTVSADGLNRFILQNGRLPECYSP 113
 DB 90 YSRNRYT-----FMDKRKSLYGKVFKNITIGTPIITSTDAEVKRVYLNQNGNFPVAPY 143
 QY 114 RSIGGILGKMSLVLDGMDRMSISLNLISHARLTILIKDVERHTLFLVDSMOQNSI 173
 DB 144 KSTIELLENSILSINGPHQRHLHTLIGAFRLSPHLKDRITRDIEASVLTLASWAQDPL 203
 QY 174 FSADEAKKFTFNIMAKIMSDGEEETEQKKKYVFMKGVSAPLNLPGTAVKALQ 233
 DB 204 VHVDEIKKMTFELIKVLMSTSPG-EDMNLKLEFEFELIGLICIPKPGTRLYSK 262
 QY 224 SRATLKEFERKMERKLDIKEEDQEEVEYKT-----EDEAENKSDHVRKQRTDDL 287
 DB 263 AKERLIKVKVVEERQVAMTTSPANDVYVLLRDGDSKQSPDSFVSGK----- 315
 QY 288 GWLKHSLMSTEQILDLISLFLGHESSVALAIAIFLQACPKAVELEBEHELEARA 347
 DB 316 -----IVEMMIGPEETPTAMTLAVKFLSDNPVALAKLEVENMKRR 358
 QY 348 KKEGESELMNDYKKMFTQCVINETRLGNVYRFLHRKLDKDRYGYIPSGKVL 407
 DB 359 KLEIGE-EYKWTDMSTLSTFQNVINETLRMANITLNGVRKALKDLEIKGYILPGWCYLA 417
 QY 408 VISAHLNDSRYDQNLNFPWRMOQNGASSSGSFSTGNNMYPGCGPRLCAGSEL 467
 DB 418 SFISVHMDIYDNPYQDPWRMDRINGSANSIC-----FTPPGGQRLCPGLSEL 468
 QY 468 AKLEMAVFHHLVLFKFMELADDDPAFPVDPNGLPIVNSRI 512
 DB 469 SKLEISIFLHLVTRYSW-TAEDEIVSFPTVKMKRRLPIVATV 512
 RESULT 6
 ID Q9M066 PRELIMINARY; PRT; 457 AA.
 AC Q9M066;
 DT 01-OCT-2000 (TREMBlrel. 15, Created)
 DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)
 DT 01-MAR-2001 (TREMBlrel. 16, Last annotation update)
 DE CYTOCHROME P450 LIKE PROTEIN.
 GN A14G36380.
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
 OC Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;
 OC Brassicales; Brassicaceae; Arabidopsis.
 OX NCBI_TaxID=3702;
 RN [1]
 RP PARTIAL SEQUENCE FROM N.A.
 RA Hilbert H., Braun M., Holzer E., Brandt A., Duesterhoeft A.,
 RA Mewes H.W., Lemcke K., Meyer K.F.X.;
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RA EU Arabidopsis sequencing project;
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
 CC -1- CATALYTIC ACTIVITY: RH + REDUCED FLAVOPROTEIN + O(2) = ROH +
 CC OXIDIZED FLAVOPROTEIN + H(2)O (BY SIMILARITY).
 CC -1- SUBCELLULAR LOCATION: MEMBRANE-BOUND. ENDOPLASMIC RETICULUM (BY

CC SIMILARITY).
 CC -1- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.
 DR EMBL: AL161589; CAB80304.1; -
 DR InterPro: IPR001105; -
 DR InterPro: IPR001128; -
 DR Pfam: PF00067; P450; 1.
 DR PRINTS: PR00385; P450.
 DR PROSITE: PS00086; CYTOCHROME_P450; UNKNOWN_1.
 DR PROSITE: PS00037; MYB_1; UNKNOWN_1.
 KW Electron transport; Endoplasmic reticulum; Heme; Membrane; Microsome;
 KM Monooxygenase; Oxidoreductase.
 SQ SEQUENCE 457 AA; 51607 MW; BD6A148E4EACD2 CRC64;

Query Match 32.1%; Score 860; DB 10; Length 457;
 Best Local Similarity 36.8%; Pred. No. 2,4e-53;
 Matches 179; Conservative 93; Mismatches 160; Indels 54; Gaps 8;

QY 39 LPPEKSGMFLGFTGYL-----KPYATTLGDMQOQVSKYKIRSNLFGPEPTVSA 92
 DB 2 IPNLSLGPVIGETLNFIFACGYSSRPV-----FMDKRKSLYGKVFKNITIGTPIIIST 55
 QY 93 DAGNRLFTLQEGRLFECSPRSIGILGKMSLVLDGMDRMSISLNLISHARLTIL 152
 DB 56 DAEVKKVVLQNHGTFVPAYPKSTIELLENSILSINGPHQRHLHTLIGAFRLSPHLKDR 115
 QY 153 LKDVERTLFLVDSMOQNSIFSADEAKKFTFNIMAKIMSDGEEETEQKKKYVTE 212
 DB 116 ITROEASVLTLASWAQDPLVHVODEIKKMTFELIKVLMSTSPG-EDMNLKLEFEF 174
 QY 213 MKGVASAPLNLPGTAVKALOSRATILKFERKMERKLDIKEEDQEEVEYKT-----E 266
 DB 175 IKGLICIPKPGTRLYSKAKERLIKVKVVEERQVAMTTSPANDVYVLLRDGCD 234
 QY 267 DEAEKMSDHRKQRTDDLLGWLKHNSLSEQILDLISLFLGHESSVALAIAIFF 326
 DB 235 SEKQSPDSFVSGK-----IVEMMIGPEETPTAMTLAVKFLSDNPVALAKLEVENMKRR 270
 QY 327 LQACPKAVELEBEHELEARA-----LQACPKAVELEBEHELEARA 386
 DB 271 LSDNPVALAKLEVENMKRRKLEIGE-EYKWTDMSTLSTFQNVINETLRMANITLNGVR 329
 QY 387 KALDVRKGYDIPSGKVLVISAHLNDSRYDQNLNFPWRMOQNGASSSGSF 446
 DB 330 KALKDVEIKGYLIPGWCYLAISFVHMDIYDNPYQDPWRMDRINGSANSIC----- 385
 QY 447 TWGNMYPGCGPRLCAGSELAKLEMAVFHHLVLFKFMELADDDPAFPVDPNGLP 506
 DB 386 -----FTPPGGQRLCPGLSELISIFLHLVTRYSW-TAEDEIVSFPTVKMKRRLP 439
 QY 507 IRVATV 512
 DB 440 IRVATV 445
 RESULT 7
 ID Q9LN73 PRELIMINARY; PRT; 478 AA.
 AC Q9LN73;
 DT 01-OCT-2000 (TREMBlrel. 15, Created)
 DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)
 DT 01-MAR-2001 (TREMBlrel. 16, Last annotation update)
 DE T12C24.27.
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
 OC Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;
 OC Brassicales; Brassicaceae; Arabidopsis.
 OX NCBI_TaxID=3702;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Chao Q., Brooks S., Buehler E., Johnson-Hopson C., Khan S., Kim C.,
 RA Shinn P., Altafi H., Bai Q., Chin C., Chiu J., Choi L., Conn L.,
 RA Conway A., Gonzales A., Hansen N., Howing B., Koo T., Lam B., Lee J.,

RA Lenz C., Li J., Liu A., Liu K., Liu S., Mukharsky N., Nguyen M.,
 RA Palm C., Pham P., Sakano H., Schwartz J., Southwick A., Thaverl A.,
 RA Toriumi M., Vaysberg M., Yu G., Federspiel N.A., Theologis A.,
 RA Ecker J.R.,
 RT "Genomic sequence for Arabidopsis thaliana BAC T12C24 from chromosome
 RT 1."
 RN Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
 RP SEQUENCE FROM N.A.
 RA Ecker J.R.,
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
 RN SEQUENCE FROM N.A.
 RP Ecker J.R.,
 RL Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.
 RN [4]
 RA SEQUENCE FROM N.A.
 RA Cheuk R., Shin P., Brooks S., Buehler E., Chao Q., Johnson-Hopson C.,
 RA Khan S., Kim C., Altafi H., Bei B., Chin C., Chlou J., Choi E.,
 RA Conn L., Conway A., Gonzalez A., Hansen N., Howing B., Koo T., Lam B.,
 RA Lee J., Lenz C., Li J., Liu A., Liu J., Liu S., Mukharsky N.,
 RA Nguyen M., Palm C., Pham P., Sakano H., Schwartz J., Southwick A.,
 RA Theologis A., Ecker J.,
 RA Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.
 CC -1- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.
 RL Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.
 CC EMBL: AC025417; AAF88087.1;
 DR InterPro: IPR001128;
 DR Pfam: PF00067; P450.1;
 DR PRINTS: PRO0385; P450.
 DR PROSITE: PS00086; CYTOCHROME_P450; UNKNOWN_1.
 DR Heme: Monooxygenase; Oxidoreductase.
 SR SEQUENCE 478 AA; 55057 MW; F67A9B9F2ED28558 CRC64;

Query Match 30.7%; Score 822; DB 10; Length 478;
 Best Local Similarity 35.6%; Pred. No. 1.3e-50;
 Matches 181; Conservative 95; Mismatches 187; Indels 46; Gaps 8;

QY 12 LLLPSSLILFLILKRRNRKTRFNLPCKSGMPFLGTYGYKPYATTATLGDPMOQH 71
 DB 4 LLIWSSLILSITHVWYSWRNPKCGKLPDGSMPGLGESIOFPKPTSDIPPIKER 63
 QY 72 VSKYGIYRSNLFGEPTVSADAGLNRFILNDEGRLEFCSYPRISGILGKMSMLVLGD 131
 DB 64 VKKGPRIKTLNVRPVVSTADSLSTFVNQEGRCFQSWTPDTTHLFGKKNWSLHGF 123
 QY 132 MHRDRSISLNFSLHARLRTLLKDERHTLFVLDSMOONSIFSAODEAKKFTFMMAKH 191
 DB 124 MYKYLKNVLLTFGHDLAK -MLPQVEMTANKRLELWSNODSVELKQDASTIIPDLTAKK 182
 QY 192 IMSDGEELFQKKEVTFMKGVSAPLNPGTAHYKALQSRATILKFERKKEERKL 251
 DB 183 LISHP -DKSEENLRANVAFIOGLISFPFPIPGTAHYKCLQGRKAKMKMLRMQERR- 240
 QY 252 DIKEEDQEEETKDEAKMSKSDHYRKQRD -DDLLGWVLKHSNLSLEQI-LDLIISL 308
 DB 241 -----EMPRKNPSDFEYVIEEIOKEGTITLLEALDLMLVL 277
 QY 309 LFAGHETSSVALAIFLQACPKAVEELREHELIARAKELGSELSMDYKMDFTQ 368
 DB 278 LFASSETTSLATLAKFLSDDEVLKRLTEHEHTLR -NRDADASGLTWETKMTYTF 336
 QY 369 CVINETLGLNVVRLHAKALKDVRK-----GYDIPSGKVLPLVISAVHLDSRYDP 422
 DB 337 QFINETALAINVPAIFKRLALDIKFEVNDVTDTITPAGMAVWVCPVAVHINPMKYDP 396
 QY 423 NLFNMRQOONNGASSSGSSEFTWGNVMPFGGPRLCAGSELAKLEMAVFIHVLK 482
 DB 397 LVFNPSRME-----GSKVTNASHKHFNAFGGMRPCVGTDTTKLOMAFLHSLVTK 446
 QY 483 FNMELAEEDQPFAPFVDFPNGLPIRVSR 511

DB 447 YRMEIRKGNITRTPGLQFPNGYHVKLHK 475
 RESULT 8
 ID Q9LY89 PRELIMINARY; PRT; 382 AA.
 AC Q9LY89;
 DT 01-OCT-2000 (Tremblrel. 15, Created)
 DT 01-OCT-2000 (Tremblrel. 15, Last sequence update)
 DT 01-MAR-2001 (Tremblrel. 16, Last annotation update)
 DE HYPOTHETICAL 43.9 KDA PROTEIN.
 GN F18022.190.
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
 OC Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;
 OC Brassicales; Brassicaceae; Arabidopsis.
 OX NCBI_TaxID=3702;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Beyer M., Murphy G., Ridley P., Hudson S., Bancroft I., Mewes H.W.,
 RA Rudd S., Lemcke K., Mayer K.F.X.,
 RL Submitted (APR-2000) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RA EU Arabidopsis sequencing project:
 RL Submitted (APR-2000) to the EMBL/GenBank/DBJ databases.
 CC -1- CATALYTIC ACTIVITY: RH + REDUCED FLAVOPROTEIN + O(2) = ROH +
 CC -1- OXIDIZED FLAVOPROTEIN + H(2)O (BY SIMILARITY).
 CC -1- SUBCELLULAR LOCATION: MEMBRANE-BOUND. ENDOPLASMIC RETICULUM (BY
 CC SIMILARITY).
 DR EMBL: AL163817; CAB87779.1;
 DR InterPro: IPR001128;
 DR Pfam: PF00067; P450.1;
 DR PROSITE: PS00086; CYTOCHROME_P450; UNKNOWN_1.
 DR Electron transport; Endoplasmic reticulum; Hypothetical protein;
 KW Membrane; Microsome.
 SR SEQUENCE 382 AA; 43888 MW; 1BC5685AB73BAE30 CRC64;

Query Match 30.4%; Score 815; DB 10; Length 382;
 Best Local Similarity 34.2%; Pred. No. 2.9e-50;
 Matches 159; Conservative 87; Mismatches 127; Indels 92; Gaps 4;

QY 45 GMPFLGTYGYKPYATTATLGDPMOQHSKYGIYRSNLFGEPTVSADAGLNRFILN 104
 DB 2 GMPFLGTYGYKPYATTATLGDPMOQHSKYGIYRSNLFGEPTVSADAGLNRFILN 61
 QY 105 GLPFCYSPRISGILGKMSMLVLGDHMRDRSISLNFSLHARLRTLLKDERHTLFV 164
 DB 62 GKLFISDPKAMHDLTGKYSLLATGELIHRKLNVIISFINITKSKPDLFCAENLSISI 121
 QY 165 LDSMOONSIFSAODEAKKFTFMMAKHIMSMDPGEETLQKKEVTFMKGVSAPLN 224
 DB 122 LKSWNREVEERHKVKTFTLSVYNQLLSIKPEDPARLYVLDDFLSYWKGFIISDPLP 181
 QY 225 GTAHYKALQSRATILKFERKKEERKLDIKEEDQEEETKDEAKMSKSDHYRKQRD 284
 DB 182 GTGYNAIKVRSNRIHONALIEDMNNAIREDFLDSITSNED----- 224
 QY 285 DLGWLKHSNLSLEQIIDLILSLFAGHETSSVALAIFLQACPKAVEELREHELEI 344
 DB 225 -----BEHNAI 230
 QY 345 ARAKELGESELNMDYKMDFTQCVINETLGLNVVRLHAKALKDVRKYGYDIPSGK 404
 DB 231 -RAKKGDEL-LNMDYQKMEFTQCVISEALRCGNIVTVHRRKATHDIKFNEYVIPKWK 288
 QY 405 VLPVISAHLDSRYDPNLFPMRQOONNGASSSGSSEFTWGNVMPFGGPRLCAG 464
 DB 289 VPIFTAAHLDSLHNEPFEFNPMTKTT-----AFQGVAVCPG 329
 QY 465 SELAKLEMAVFIHVLKFNMLEAEDQPFAPFVDFPNGLPIRV 509

Db 330 GELCKLQIAFLHHLVLSYRWKIKSDEMPLAHYVEFKGMLLEI 374

RESULT 9 PRELIMINARY; PRT: 464 AA.

09LIC5

AC 09LIC5; PRELIMINARY; PRT: 464 AA.

DT 01-OCT-2000 (Tremblrel. 15, Created)

DT 01-OCT-2000 (Tremblrel. 15, Last sequence update)

DT 01-MAR-2001 (Tremblrel. 16, Last annotation update)

DE CYTOCHROME P450-LIKE PROTEIN.

OS Arabidopsis thaliana (Mouse-ear cress).

OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;

OC Brassicales; Brassicaceae; Arabidopsis.

NCBI_Taxid=3702;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN-COLUMBIA; Kaneo T., Kato T., Nakamura Y., Asamizu E., Tabata S.;

RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.

RN [2]

RP SEQUENCE FROM N.A.

RC STRAIN-COLUMBIA; Kaneo T., Kato T., Nakamura Y., Asamizu E., Tabata S.;

RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.

RN [2]

RP SEQUENCE FROM N.A.

RC STRAIN-COLUMBIA; Kaneo T., Kato T., Nakamura Y., Asamizu E., Tabata S.;

RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.

RN [2]

RP SEQUENCE FROM N.A.

RC STRAIN-COLUMBIA; Kaneo T., Kato T., Nakamura Y., Asamizu E., Tabata S.;

RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.

RN [2]

RP SEQUENCE FROM N.A.

RC STRAIN-COLUMBIA; Kaneo T., Kato T., Nakamura Y., Asamizu E., Tabata S.;

RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.

RN [2]

RP SEQUENCE FROM N.A.

RC STRAIN-COLUMBIA; Kaneo T., Kato T., Nakamura Y., Asamizu E., Tabata S.;

RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.

RN [2]

RP SEQUENCE FROM N.A.

RC STRAIN-COLUMBIA; Kaneo T., Kato T., Nakamura Y., Asamizu E., Tabata S.;

RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.

RN [2]

RP SEQUENCE FROM N.A.

RC STRAIN-COLUMBIA; Kaneo T., Kato T., Nakamura Y., Asamizu E., Tabata S.;

RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.

RN [2]

RP SEQUENCE FROM N.A.

RC STRAIN-COLUMBIA; Kaneo T., Kato T., Nakamura Y., Asamizu E., Tabata S.;

RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.

RN [2]

RP SEQUENCE FROM N.A.

RC STRAIN-COLUMBIA; Kaneo T., Kato T., Nakamura Y., Asamizu E., Tabata S.;

RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.

RN [2]

RP SEQUENCE FROM N.A.

RC STRAIN-COLUMBIA; Kaneo T., Kato T., Nakamura Y., Asamizu E., Tabata S.;

RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.

RN [2]

RP SEQUENCE FROM N.A.

Query Match 29.4%; Score 789; DB 10; Length 464;

Best Local Similarity 36.5%; Pred. No. 2,7e-48;

Matches 163; Conservative 101; Mismatches 145; Indels 38; Gaps 8;

40 PRGSGMPFLGEGTIGLYKPYATTLGDMOQVSKYKTIYSNLFGEPTIYSADAGLNRF 99

52 PHGSLGWVIGETIEFVSAYSADRPESFMDKRRMLYGRVFSHFPGATIVSTDAEVNRA 111

100 ILONEGRLEFECSPRISIGILGKMSMLVGDHMRMSISLNFSLHARLRTLLKDYER 159

112 VLQSDSTAFVFFYKTYRELKMGKSSILLINGSLSRRRGLVGLSPKSLLAQIVRDHMK 171

160 HTLEVLDSMOQNSIFSADKAKFTFLMAKHMISMDGEEETOLKKEYTFMKGVVSA 219

172 FLSESMDLMSDQVLLQDVSKTVAFLAKALISVEKG-EDLEELKREFENFISGLMSL 230

220 PLNPGRAIHALDSKATILKFIKRMKEKLDIKEDQEEVEKTEDEAM--SKSDH 276

231 PINPGQLHSLAKNMVQYVERIIEGKIR--KTKNKEDDVIADVDVLLKDSSEH 288

277 VRKQRTDDLLGWLKHSNLSSTEOILDLISLFAHETSVAIALAIFLQACPKAVEE 336

289 -----LTH-NLIANNMIDMT---PGHDSVFLITLLVAKFLSDSPALNL 329

337 LREHLEIARAKKELGSELNMDYKMDFTQCYINELRLGNVVRFLHRKALDKVRYK 396

330 LFEENMML-KSLKELTGEPVLYMNDYLSLPTQKYITELRLKGNVILIGVYRKAMDVEIKG 388

397 YDISGKAVLPVIAVHLDNSRITDQPNLEFNPWQOONNGASSGSGSFSFGNNYMPFG 456

389 YVIFKGMCFIAYLRSVHLDKLYESPYKFNPMWQERDMNTS-----FSPFG 436

QY 457 GPRLCAGSELEKEMAVFIHHLVLE 483

Db 437 GGORLCPEGLDARLETSYFLHHLVTRF 463

RESULT 10

09FMA5

AC 09FMA5; PRELIMINARY; PRT: 465 AA.

DT 01-MAR-2001 (Tremblrel. 16, Created)

DT 01-MAR-2001 (Tremblrel. 16, Last sequence update)

DT 01-MAR-2001 (Tremblrel. 16, Last annotation update)

DE CYTOCHROME P450.

OS Arabidopsis thaliana (Mouse-ear cress).

OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;

OC Brassicales; Brassicaceae; Arabidopsis.

NCBI_Taxid=3702;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN-COLUMBIA; Kaneo T., Kato T., Nakamura Y., Asamizu E., Miyajima N.,

RA Sato S., Kaneo T., Kotani H., Nakamura Y., Asamizu E., Miyajima N.,

RA Tabata S.;

RT "Structural analysis of Arabidopsis thaliana chromosome 5. IV. physically assigned pl and TAC clones."

RT DNA Res. 5:41-54(1998).

RL EMBL: AB009048; BAB08653.1;

DR EMBL: AB009048; BAB08653.1;

SO SEQUENCE 465 AA; 53767 MW; 0C00459C9C866DIF CRC64;

Query Match 27.1%; Score 726; DB 10; Length 465;

Best Local Similarity 32.3%; Pred. No. 8.4e-44;

Matches 166; Conservative 91; Mismatches 181; Indels 76; Gaps 10;

12 LLLPSSLTLLFLILKRRN--RTRFNLPPGSGMPFLGEGTIGLYKPYATTLGDMO 69

11 LLIVSLCSALL-----RKNQMTYKNGLEPRGTMGPIFEBTEFLKQGP-----NEMR 59

70 QHVSRYGKIVRSNLFGEPTIYSADAGLNRLFLONEGRLEFECSPRISIGILGKMSMLV 129

60 NQRLRYSGFSEKSHLGLCTPLISMSEVNRITLKNSKGLVGPQSMIDLGTGNMAVH 119

130 GDMRDRKRSISLNFSLHARLRTLLKDYERHTLVLDMSMOQNSIFSADKAKFTFL 185

120 GSHRLKMGSLSLISSTMDMDHILPKVDHFMRSYLDQWNELEVIDIQKTHMAFLSSL 179

186 -----NLMAKHMISMDGEEETOLKKEYTFMKGVVSAFLNPGTAHAKALQSRATILK 240

180 TQINGNLKRPV-----EEFKTAFKLVYGTLSVPLDLEGTNRGCIQARNNIDR 229

241 FIERKMEERKLDIKEDQEEVEKTEDEAMSKSDHVRKQRTDDLLGWLKHSN---L 296

230 LLRLMQRER-----DSGE-----TFIDMLGYLKKMKGNRYPL 262

297 STEQILDILSLFAHETSVAIALAIFLQACPKAVEELREHLEIARAKKELGSEL 356

263 TDEIRQOVVITILSGYETVSTSMALKYLDHPKALQELRAHLEAFERRKRO--DEPL 320

357 NMDYKMDFTQCYINELRLGNVVRFLHRKALDKVRYKMDFTQCYINELRLGNVVRFL 416

321 GLEVYKMKFTRAVIYETSLRATLVNGVLRKTTDLDELNGILKGRIRIYTTREINDA 380

417 SRYPDNLNFMWQOONNGASSGSGSFSFGNNYMPFGGPRLCAGSELEKEMAVFI 476

381 NLVEDPLIFNPMWQOONNGASSGSGSFSFGNNYMPFGGPRLCAGSELEKEMAVFI 429

477 HHVLEKFMELAEEDDOPFAFPVDFPGLPIRVS 510

430 HYFTVIRMEIIGDELMVFPFVPAKGFHLRIS 463

RESULT 13
065624 PRELIMINARY: PRT: 457 AA.
AC 065624;
DT 01-AUG-1998 (TREMblrel. 07, Created)
DT 01-AUG-1998 (TREMblrel. 07, Last sequence update)
DT 01-MAR-2001 (TREMblrel. 16, Last annotation update)
DE CYTOCHROME P450.
CN T18B16.200 OR AT4G19230.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
OC Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;
OC Brassicales; Brassicaceae; Arabidopsis.
NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RA Bevan M., Benes V., Rechmann S., Borkova D., Ansoerge W., Bancroft I.,
RA Mewes H.W., Mayer K., Schueller C.;
RL Submitted (JAN-1998) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA Benes V., Rechmann S., Borkova D., Ansoerge W., Mewes H.W., Lemcke K.,
RA Mayer K.F.X.;
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE OF 131-457 FROM N.A.
RA Van Der Schueren J., Chuang Y.J., Aert R., Defoor E., Robben J.,
RA Volckert G., Mewes H.W., Lemcke K., Mayer K.F.X.;
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
RN [4]
RP SEQUENCE FROM N.A.
RA EU Arabidopsis sequencing project;
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.
DR EMBL; AL021687; CAAL6713.1; -
DR EMBL; AL161550; CAB78925.1; -
DR HSSP; P33006; ICPT.
DR InterPro; IPR001128; -
DR Pfam; PF00067; P450; 1.
DR PRINTS; PR00385; P450.
DR PROSITE; PS00086; CYTOCHROME_P450; UNKNOWN_1.
KW Heme; Monooxygenase; Oxidoreductase.
SQ SEQUENCE 457 AA; 52436 MW; 6105FB7C1B1E4F07 CRC64;

Query Match 23.6%; Score 632; DB 10; Length 457;
Best Local Similarity 31.88; Pred. No. 4e-37;
Matches 151; Conservative 92; Mismatches 176; Indels 56; Gaps 12;
QY 13 LLLPSSLILFLILKRRNRKTRFNLPPGKSGWPLGEGTIGYLRATYATTLGDFMOQH 72
DB 10 LAGSLFLYFLRCLISORFGSSKLPPLPGTMGMPYGET---FQLYSQDP-NVFPQSKQ 65
QY 73 SKYGIYNSNLFGEPTIYSAAGLNRLFLONEGRLEFCSYSTRSIGILGKMSMLVNDM 132
DB 66 KRYGSVFTHVLGQPCWMISSPEAKFVLVTKSHLFKTFPASKERMGKQAIFFHQGDY 125
QY 133 HRDRSTISLNFSLHARLTILKIDVERHTLFVLDWMQONSIFSAODEAKKFTFNLMAHI 192
DB 126 HAKLRKLVLRAMPESIN-NVPDIESTAQDSLRSW-EGTIMINTYOEKKTITFENVALIS 183
QY 193 MSMDPGEET---EOLKEVYTFMKGVASAPLNLPGTAYHAKLOSRAITLFIERKMEER 249
DB 184 F---GKREVLRYRDLKRCYILIEKGYMSMVPNLPGTLFHKSMARKELSDILARILSER 239
QY 250 KLIDKEEDQEEVEKTEDEAMSKSDHYRKORTDDDLG-NVLKHSNISTEQIIDLILSL 308
DB 240 R-----QNGSSH-----NDLGSFSGKEELTDEQIDADNITIGV 272
QY 309 LLAGHETSSVALAIFLQACPKAVELREHLEIARAKKELGESLNMWDYKMDPTQ 368
DB 273 IFARDTITASWMLKYLAEINPVLEAVTEEQMAI-RKDEEGES-LTWGDTKRMPLTS 330

QY 369 CVINETLRGNVREPLHRRKALKDYRKCYDIPSGWKVLPVISAVHLNDSRYDOPNLFPMW 428
DB 331 RVIGETLKVASILSTTFREAVEDVEYEGTLLPKGKVLPLRNIHSHADITSNECKFPDS 390
QY 429 RMQOONNCGSSSGSFSTGNNVMPFEGGPRCLAGSELAKLEMAVFLHNLVLR 483
DB 391 RFE-----VAPKNTFMFPCNGTHSCGNELAKLEMSIMIHLLTXY 432
RESULT 14
Q9F138 PRELIMINARY: PRT: 518 AA.
AC Q9F138;
DT 01-MAR-2001 (TREMblrel. 16, Created)
DT 01-MAR-2001 (TREMblrel. 16, Last sequence update)
DT 01-MAR-2001 (TREMblrel. 16, Last annotation update)
DE CYTOCHROME P450-LIKE PROTEIN.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
OC Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;
OC Brassicales; Brassicaceae; Arabidopsis.
NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RA STRAIN-COLUMBIA;
RX MEDLINE=99397451; PubMed=10470850;
RA Kaneo T., Katoh T., Sato S., Nakamura Y., Asamizu E., Kotani H.,
RA Miyajima N., Tabata S.;
RT "Structural analysis of Arabidopsis thaliana chromosome 5. IX.
RT Sequence features of the regions of 1,011,550 bp covered by seventeen
RT P1 and TAC clones."
RL DNA Res. 6:183-195(1999).
DR EMBL; AB017064; BAB11064.1; -
SQ SEQUENCE 518 AA; 58877 MW; CCA3F74F4301563E CRC64;

Query Match 23.5%; Score 631; DB 10; Length 518;
Best Local Similarity 30.5%; Pred. No. 5.7e-37;
Matches 152; Conservative 101; Mismatches 190; Indels 56; Gaps 9;
QY 27 LKRRNRKTRFNLPPGKSGWPLGEGTIGYLRATYATTLGDFMOQHSYKGIYNSNLFGE 86
DB 64 LYRWSNPKCNCKLPPGSGWPLIGETCFEPHGLYEISPPVKRMKLKGLFRFINIGS 123
QY 87 PTIYSAOAGLNRLFLONEGRLEFCSYSTRSIGILGKMSMLVNDMHRDMSISLNFSLH 146
DB 124 NTVVLTEPDITFEYFRQDNKSFVSYPAYVKPRGKENVFLKHGNIHNVKQISLQHLGS 183
QY 147 ARLTILKIDVERHTLFVLDWMQONSIFSAODEAKKFTF-NLMAKHMSMDPGEETEOL 205
DB 184 EALKKMGELDRVTEHLRSKANQSGFPAKAVESVIMAHLPKTIISNLKPEQAT--L 241
QY 206 KKEYVTFMKGVASAPLNLPG-TAYHAKLOSRAITLFIERKMEERK-----LDIKE 255
DB 242 VDNIMALGSEWFOPLKLTLLISITKYFIARVALQYIKVDVTRRKASREMGDFLDMV 301
QY 256 EDOEEVEKTEDEAMSKSDHYRKORTDDDLGWLKHSNISTEQIIDLILSLFAGHET 315
DB 302 EGEKEEDYENES-----AINLITAILVAKES 330
QY 316 SVAVALAIFLQACPKAVELREHLEIARAKKELGESLNMWDYK-KMDPTQCVINET 374
DB 331 TTSVTSIAIKFLAEHAKALAEKREHAILLONRNGK-AQSVWEERYRQMTFMVNIET 389
QY 375 LRLGNVREPLHRRKALKDYRKCYDIPSGWKVLPVISAVHLNDSRYDOPNLFPMRQOON 434
DB 390 LRMANMAPIMYRKAVNDVEIKGYTIPAGVIAPPAVHFMDAITEPLEFNPWRBEKE 449
QY 435 NGASSGSGSFSTGNNVMPFEGGPRCLAGSELAKLEMAVFLHNLVLFKNELADDPF 494
DB 450 ---LRSGSKTF-----WFGGVRCVGAEFARLQISTIFHLVTTTDSLQAESFT 499

